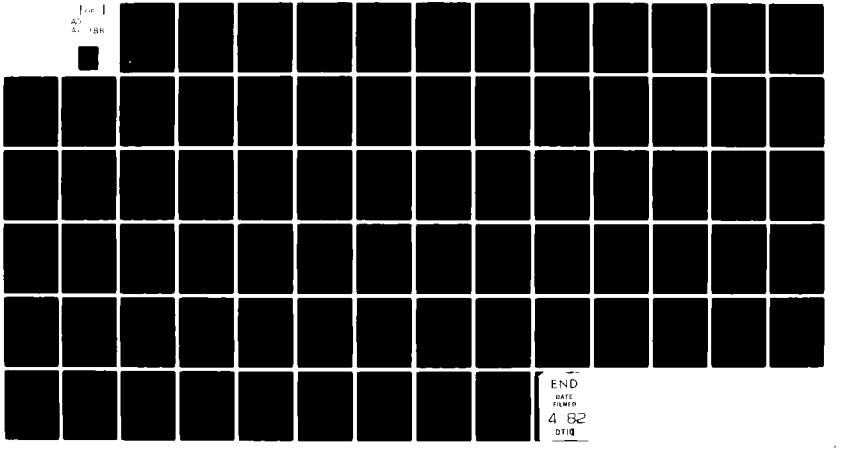
AD-A111 188 ARIZONA UNIV TUCSON DEPT OF PSYCHOLOGY F/G 5/10
STRATEGIES FOR ABSTRACTING MAIN IDEAS FROM SIMPLE TECHNICAL PRO--ETC(U)
NOV 81 D E KIERAS, S BOVAIR N00014-78-C-0509
UNCLASSIFIED UARZ/DP/TR-81/9 NL



END
DATE FILMED
4 82
DTIC

AD A111188

Strategies for Abstracting Main Ideas From Simple Technical Prose

David E. Kieras
and
Susan Bovair
University of Arizona

DTIC
ELECTE
FEB 2 2 1982
A

Technical Report No. UARZ/DP/TR-81/9
November 10, 1981

This research was supported by the Personnel and Training Research Programs, Office of Naval Research, under Contract Number N00014-78-C-0509, Contract Authority Identification Number NR 157-423. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for Public Release; Distribution Unlimited.

DTIC FILE COPY

82 02 22 048

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER UARZ/DP/TR-81/9	2. GOVT ACCESSION NO. AD-A111 188	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Strategies for Abstracting Main Ideas from Simple Technical Prose		5. TYPE OF REPORT & PERIOD COVERED Technical Report Nov 10, 1981 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David E. Kieras and Susan Bovair		8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0509
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Arizona Tucson, AZ 85721		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N; RR 042-06 RR 042-06-02; NR 157-423
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE 11 November 1981 13. NUMBER OF PAGES 76
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Reading, Comprehension, Abstraction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents detailed results on performance in a comprehension task in which the reader must devise a brief statement of the main idea of short technical passages. The passage structure consisted of a generalization followed by several examples, and appeared either with or without an initial "topic sentence" stating the generalization. Data on response content, reading time, ratings of importance of passage sentences, and "think aloud" protocols were

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

collected. The results suggest that most readers use a simple strategy tailored to the generalization structure of the passages. This strategy reflects both a reliance on the surface structure of the passage, such as what is first mentioned, and use of a moderate, but not complete, understanding of the actual passage content. Some subjects were found to be defective in their strategy; the most clearly defined defect consisting of a failure to recognize the generalization nature of the main idea. The prevalent strategy was represented in the form of a computer simulation using production systems and propositional memory structures. The simulation was found to be reasonably accurate in several respects. Especially interesting is the fact that relatively little general knowledge is needed by the model.

Abstract

This report presents detailed results on performance in a comprehension task in which the reader must devise a brief statement of the main idea of short technical passages. The passage structure consisted of a generalization followed by several examples, and appeared either with or without an initial "topic sentence" stating the generalization. Data on response content, reading time, ratings of importance of passage sentences, and "think aloud" protocols were collected. The results suggest that most readers use a simple strategy tailored to the generalization structure of the passages. This strategy reflects both a reliance on the surface structure of the passage, such as what is first mentioned, and use of a moderate, but not complete, understanding of the actual passage content. Some subjects were found to be defective in their strategy; the most clearly defined defect consisting of a failure to recognize the generalization nature of the main idea. The prevalent strategy was represented in the form of a computer simulation using production systems and propositional memory structures. The simulation was found to be reasonably accurate in several respects. Especially interesting is the fact that relatively little general knowledge is needed by the model.

A

Strategies for Abstracting Main Ideas from Simple Technical Prose

David E. Kieras
and
Susan Bovair

University of Arizona

This paper is concerned with how people abstract the main idea from a piece of technical prose in the main idea task, in which people read a paragraph-length technical passage, and then make up a brief, one-sentence statement of the main idea of the passage. The most useful theoretical formulation for this task is the macrostructure theory of comprehension developed by Kintsch and van Dijk (1978). This theory was devised to explain prose memory phenomena. Basically, the input text is first processed at a low level, resulting in microstructure propositions which express essentially the immediate content of the passage. Then, macro-processes, using general knowledge, condense the microstructure down to a relatively few macropropositions which express the gist, or important content, of the passage. These macropropositions are then put into memory. When it is time to recall the passage content, the macropropositions are retrieved, and then general knowledge is used to reconstruct some of the micropropositions, which of course may be rather different from those originally presented. These are then recalled, resulting in recall which has the same gist as the original, but will usually be highly paraphrased and condensed.

In the main idea task, subjects are expressing in their main idea statement the central part of their macrostructure for the passage. Since little or no memory encoding and retrieval is involved, this task yields direct information on how readers derive the passage macrostructure. But in contrast to recall paradigms, this task conveys little information on the memory phenomena associated with passage macrostructure.

The rules for deriving macrostructure have been proposed for some time (van Dijk, 1977a,b; 1980). These rules express how a set of microstructure propositions can be replaced with a smaller number of macropropositions, based on the semantic content, both explicit and inferred, of the passage. But there has been little direct study of the operation of these rules. One immediate complication is that surface-level aspects of the passage, as well as the semantic content, appear to be important in determining the passage macrostructure. Work by Kieras (1978,1980,1981), Kozminsky (1977), Clements (1979), van Dijk (1979), and Perfetti and Goldman (1974, 1975) has focused on specific

aspects of how the topic or main idea of a passage is marked or signalled to the reader. Initial mention, in the form of a traditional topic sentence, is one cue (Kieras, 1980), a title is another (Kozminsky, 1977), and more subtle markers, such as topic-comment assignment at the sentence level, are another (Kieras, 1981; van Dijk, 1979). Hence, an adequate theory of macro-processes must explain not just the use of semantic content in defining the main content, but also the use of these surface-level features.

This paper attempts to present a detailed examination of a relatively simple form of macrostructure building. The focus is on the strategies used by readers to abstract the main idea. Normal readers' strategies are suggested, based on several kinds of experimental data obtained from readers, along with some results on readers who have defective strategies. Then a simple simulation model of the macrostructure building process is presented, which uses the normal strategies.

EXPERIMENTAL RESULTS

The passages used in this work have a simple structure. They begin with a generalization, and then present several examples or instances of the generalization, with some unimportant items included as well. One set of passages has been studied very intensively in this work. Each passage appeared in two versions; In the good version, the generalization was explicitly stated in the first sentence. In the bad version, this statement was deleted, and the first sentence was identical to the good version's second sentence.

The experimental data was collected by presenting the passages to subjects one sentence at a time, in a self-paced procedure. Four sets of data were collected in three experiments. In all of these, the subjects read all of the passage sentences, and then composed a statement of the main idea. In the first study, subjects provided reading times during the reading phase, and after entering their main idea statement, they were shown each sentence again and then rated their prior pre-experimental familiarity with the content of the sentence. In the second study, subjects rated the importance of each sentence during the first reading. A qualitative rating scale was used, consisting of Central to the main idea, Related to the main idea, and Unimportant. The content of the main idea responses was compared for the two versions. In a third study, think-aloud protocols were collected in which subjects were intructed to state their current hypotheses of the main idea, and how they arrived at it, after reading each sentence.

Method

Materials

Four passages were prepared, based on those studied in Kieras (Note 1). The structure of these passages consisted of a generalization followed by several examples of the generalization, with some superfluous material included as well. Each passage appeared in two versions, a good version, in which the generalization was explicitly stated in the first sentence, and a bad version, in which the explicit statement of the generalization was missing, but all the other sentences were the same as in the good version. The passages were deliberately prepared to have a variety of sentence forms and sentence lengths, and were also prepared to vary in length, both to be more natural, and to ensure that in a sentence-at-a-time paradigm, the reader could not confidently expect the passage to be of a certain length. The four passages were intended to vary in overall familiarity of their content, based on earlier pilot work, and also to vary in the familiarity of the content of the individual sentences in each passage. Tables 1, 2, 3, and 4 show the four passages, referred to as METALS, TIMEKEEPING, INSTRUMENTS, and CARS. In each passage the first sentence, shown in brackets, was deleted to produce the bad version. The sentences are numbered starting with this good version first sentence, and these numbers will be used to refer to the individual sentences in each version. Hence, the first sentence in the bad version is Sentence 2, and the last sentence in both versions of the METALS passage is Sentence 14.

One of the products of this work is an emphatic demonstration of how each passage is unique, even though an overall similarity in structure was intended. For this reason, the actual content of each passage is important to understanding the results. The reader will find it useful at this point to read through the four passages and notice their individual content.

Design

In all three experiments the same experimental design was used. Each subject read and responded to all four passages, but saw only one of the two versions of each passage, getting two good versions and two bad versions. Which versions of the passages an individual subject saw was determined at random for pairs of subjects, so that in each consecutive pair of subjects, each passage appeared once in each version. With an even number of subjects run, each passage thus appeared an equal number of times, and an equal number of times in each version. The order of appearance of the four passages in the experiment was randomized for each

Table 1

The METALS passage

---

1. [Different cultures have used metals for different purposes.]
2. The ancient Hellenes used bronze swords.
3. The ancient Greeks used copper shields.
4. The Hellenes invaded ancient Greece before the Trojan War.
5. The bronze weapons that were used by the Hellenes could cut through the copper shields that were used by the Greeks.
6. Because the color of gold is beautiful, the Incas used gold in religious ceremonies.
7. The Incas lived in South America.
8. However, the Spaniards craved the monetary value of gold.
9. Therefore, the Spaniards conquered the Incas.
10. Because aluminum does not rust and is light, modern Western culture values aluminum.
11. Aluminum is used in camping equipment.
12. Titanium is used in warplanes and is essential for spacecraft.
13. Warplanes are extremely expensive.
14. Titanium is the brilliant white pigment in oil paints that are used by artists.

Table 2

The TIMEKEEPING passage

---

1. [Modern timekeeping devices are extremely accurate.]
2. An inexpensive quartz-crystal watch has one-second accuracy for several weeks.
3. Proper adjustment of the watch can improve the accuracy.
4. An atomic resonance clock can achieve nano-second accuracy for several years.
5. The theory of relativity predicts that tiny distortions of time would be produced on a long trip in a commercial airliner.
6. Because atomic resonance clocks are very accurate, they could measure the tiny distortions of time and confirm the theory.
7. A hydrogen maser clock has pico-second accuracy for 10 million years.
8. A hydrogen maser clock is used today by the National Bureau of Standards.

Table 3

The INSTRUMENTS passage

---

1. [Because keyboard instruments have different mechanisms, the performer can control different aspects of the sound of the instrument.]
2. The clavichord is the oldest keyboard instrument.
3. The clavichord has a small metal hammer at the end of the key.
4. When the hammer strikes the string, the string vibrates between the hammer and the bridge.
5. Since the key is in direct contact with the string, the player can control the pitch.
6. The harpsichord has a small stiff finger that plucks a string.
7. Since the finger always moves through the same distance, the performer can not control the loudness of the sound.
8. Finally, the piano has a hammer that is bounced off a string.
9. The force that is applied by the hammer depends on the force that is applied to the key.
10. This means that the performer can control the loudness of the individual notes.
11. Therefore, the piano is the most expressive instrument.

Table 4

The CARS passage

---

1. [Different automobiles are selected by people who prefer different features.]

2. Imported luxury cars are expensive and have advanced design.

3. They are owned by people who are wealthy and appreciate sophisticated cars.

4. They often have electronic fuel injection systems that are controlled by analog computers.

5. Because domestic station wagons are roomy and comfortable, they are preferred by people who have large families.

6. The original station wagons had bodies that were mostly made of wood.

7. The pickup is a small open truck that can carry a large amount of cargo and is preferred by many people who live in rural areas.

8. Since compact cars are small and have small engines, they give good gas mileage.

9. This means that people who commute like compact cars.

10. Most compact cars are made by foreign manufacturers.

11. Because gasoline was cheap, the first American compact car was a failure and caused the bankruptcy of the manufacturer.

12. Since sports cars are tiny and fast, people who enjoy driving like sports cars.

13. Until the Corvette appeared, all sports cars were imported.

---

subject.

Subjects

The subjects for the Reading Time, and Rating experiments were recruited from the student population at the University of Arizona through advertisements, and were paid $2 for participating. The numbers were 114 for the Reading Time, and 72 for the Rating experiment. The Protocol subjects were chosen differently, because it was felt to be crucial to get subjects who were certain to be highly articulate and willing to engage in the "think aloud" task. Ten subjects were individually recruited, mostly from the psychology graduate students at the University of Arizona, who were, however, unexposed to cognitive psychology and reading comprehension research. Due to the time and effort involved, these subjects were paid $5 for participation.

Procedure

Subjects were run in groups of 1-3 using a laboratory computer (Kieras, 1979). The computer presented the sentences one at a time on video terminals in a self-paced procedure, performed the randomizations, and recorded responses and reading times. The subject first read a set of instructions on how to type in responses on the terminal, followed by a brief typing practice period. Then the subject read a set of instructions for the experimental task, was checked for understanding by the experimenter, and then performed the task on a practice passage. After being checked once more, the subject then began the experimental task on the four passages. The basic procedure for all three experiments was the same, with modifications as described below for the different experiments. The computer first presented a prompting message, and then the subject tapped the space bar on the keyboard to make the first sentence appear. After reading the sentence, the subject tapped again, which made the first sentence disappear, and the next sentence appear, and so forth through the entire passage. The time each sentence was left on the screen was recorded as the reading time. After the last sentence, a prompt would appear for the subject to type in a statement of the main idea. After the subject entered the response, the prompt for the first sentence of the next passage would appear.

Reading Time Experiment. The subject read each sentence, with the reading time recorded, and then entered a main idea statement. Then the subject saw each sentence in the same passage again, and rated the how much of the information in the sentence he or she knew prior to the experiment. These familiarity ratings were performed on a 1

(knew none of it) to 7 (knew all of it) scale. Then the subject proceeded to the next passage. The instructions for the main idea statement were like those in Kieras (Note 1); the subject was to devise a short (80 characters maximum) complete sentence that stated what he or she thought was the main idea of the passage. Also included in this experiment were two other passages of a different type which were included to obtain pilot data; the results for these will not be reported, and they were not included in any subsequent experiments.

Rating Experiment. While reading each sentence, the subject rated the importance of the sentence to the main idea of the passage. After the last sentence, the subject entered a main idea statement as in the Reading Time experiment. In an attempt to get ratings information more directly comparable to the simulation model, a three-point qualitative scale was used rather than the usual 7-point quantitative scale. The subject judged the sentence as being Central (C) if it either stated the main idea, or made them change their mind about the main idea; Related (R) if it was just related to the main idea, or Unimportant (U) if the sentence was unimportant to the main idea. The session required about an hour.

Protocol Experiment. The subjects were individually run, with the experimenter present, and the subject's "thinking aloud" being tape recorded. The instructions asked subjects to read each sentence aloud, and then to state their current idea of the main idea of the passage and how that particular sentence fit in, "thinking aloud" on how they arrived at their decisions. They also thought aloud while preparing their main idea statement at the end of the passage. Although instructed to state their current main idea on each sentence, lapses were common; the experimenter attempted to prompt the subject as needed, with care taken not to influence the subject's thinking. The sessions required a full hour. Tape recorder failures made it necessary to replace 2 subjects to arrive at the desired sample of ten.

Results

The actual body of the results will consist of a passage-by-passage presentation. Here will be summarized some overall analyses and the methods used in the passage-by-passage analyses.

The reading time data from the Reading Time experiment were averaged across subjects to produce a mean reading time for each sentence in each passage. An analysis of variance was performed on the reading time data for each passage, using Sentence, Version, and Subjects as factors, with Sentence 1 being excluded since it appeared only in the good

version. These analyses showed strong sentence effects (all ps<.01), but version main effects appeared only in METALS (p<.05), with the other passages very non-significant (ps>.2). Significant (p<.01) interactions of sentence and version appeared in METALS and INSTRUMENTS and marginally in CARS (p=.08). On the whole, these analyses show that the version manipulation had some effect on reading times, but not always (cf. Kieras, 1980, 1981). To supplement the ANOVAs, individual t-tests were computed for each sentence to compare the reading times in the two versions. These specific results are presented below.

The familiarity ratings were averaged across subjects to yield a mean familiarity rating for each sentence. The primary use of this data was as a predictor variable for the reading time, as described below. But analysis of variance confirmed the desirable features of this measure, that it varied strongly between sentences (ps<.001), and not at all between versions, or in interaction between sentence or version (ps>.1). By descending order of mean familiarity of the passage sentences, the passages and their means are: CARS (5.5), METALS (4.6), INSTRUMENTS (4.0), and TIMEKEEPING (3.1).

The ratings data were tabulated to show the proportion of responses of each type (C, R, or U) on each sentence in each passage version. Individual chi-square tests were used to compare the distribution of responses for each sentence to detect version differences. These results are presented in the passage by passage analysis below.

The analysis of the protocols was difficult and time-consuming since a standard methodology is not available. The tape recordings were first transcribed verbatim, and then in two passes, were condensed using a small standardized set of descriptions shown in Table 5. This condensed description summarized the decision made by the subject concerning the status of the sentence, the status of the current main idea, and the processing involved with these decisions. For presentation here, these condensed descriptions were further condensed to show just the critical individual actions performed on each sentence. Since the subjects were very strongly different in their actions, these results will be shown for individual subjects. One caveat must be made; apparently the protocol subjects are not directly comparable to the subjects in the other experiments, in that their protocols show considerably more revising of the current main idea than is plausible for subjects in the other tasks. Apparently, they responded to the task demands by indulging in very extensive and subtle processing. The protocol results will be presented with each passage.

Table 5

Protocol Condensation

STATES		: <statement> subject states current main idea of passage.
HYPOTHESIZE	TOPIC	: <statement>
	MI	subject suggests a possible topic or main idea for the passage.
PREDICT	SENTENCE	: <statement>
	DIRECTION	subject predicts either what the next sentence will be, or the general direction of the passage.
	INFORMATION	
RECALL	MI	: <statement>
	SENTENCE	subject recalls from memory: information from earlier in the passage, a previous main idea, or an earlier sentence in the passage.
RELATES		: <statement> subject describes how the current sentence relates to the main idea.
COMMENT		: <statement> subject makes a statement about the sentence or the main idea not covered by any other verb.

The main idea responses were examined using two procedures, one gross, one fine. In the gross analysis, the responses were sorted into categories, on the basis of simple apparent similarity in content, in order to produce roughly 10 groups of responses. Each category was then described by an exemplar or a composite of exemplars. This sorting process was done blind with regard to the original version of the passage associated with the response; hence no systematic effects of the looseness of this process would be expected. After the sorting was complete, the responses were then separated by the passage version, and the number of responses in each category were counted. The distribution of responses in the two versions can be compared with an ordinary chi-square test. This analysis shows in a simple way the nature of the responses and the version differences.

The fine analysis consisted of first constructing a propositional representation of each response, using the rules presented in some detail in Bovair and Kieras (Note 2), based on Turner and Greene (Note 3) and Kintsch (1974). Two independent judges constructed these representations, which were then reconciled. A LISP program was used to tabulate the individual predicates, arguments, and propositions appearing in the responses. A "synonymization" step was then performed, in which predicates, arguments, and propositions which appeared to be similar in meaning were replaced with a single term, ensuring that minor differences in meaning and variations in the original propositional analysis of the responses would be minimized or eliminated. All these steps were done blind with regard to the version of the passage associated with the response. The responses were then separated by the original version, and the individual propositions tabulated, and their frequency of appearance counted. The number of subjects producing propositions in responses made to the two versions can be compared as follows: For each proposition, each subject can be classified as either producing the proposition, or not. The difference in proportion of producers between versions can be tested with chi-square. By making the (questionable) assumption that each proposition is independent of the others, a total comparison of the two sets for production frequencies for all of the propositions can be made by summing the individual chi-square values. Since there are many propositions that are produced by only very few subjects, the list of propositions reported and used in the comparison was truncated by including only propositions produced by at least five subjects in at least one version.

Overview of Results

There are certain recurring features in the results which can be pointed out before presenting the passage-by-passage results. One striking pattern is the treatment of the first sentence in the passage, which in the good version is Sentence 1, the explicit statement of the generalization, and in the bad version is Sentence 2. In the good version, the first sentence is uniformly recognized as being a statement of a main idea, and is rated very high in importance, read for a relatively long time, and described as a good main idea in the protocols. Many of the main idea statements for the good version essentially reproduced the first sentence. In the bad version, the first sentence may or may not be considered important; in two of the passages, it is rejected as a main idea statement; but in the other two, Sentence 2 turns out to be a satisfactory topic sentence, but the main idea based on it turns out to be inconsistent with the rest of the passage.

While reading the body of the passage, the sentences in the good version are compared to the first sentence, and are usually accepted as exemplars of the main idea, and there are relatively few revisions of the main idea. In the bad version, there are generally many revisions. Since the passages were prepared so as to be based on the good version generalization sentence, the revisions made while reading the bad version have a strong tendency to eventually arrive at the same main idea that the good version states explicitly. The main focus of the results to be presented is specifically how these effects appear in the individual passages.

The METALS Passage

Responses. Table 6 shows the distributions of responses in the categories for the two versions, which were significantly different ($X2(7)=18.128$, $p<.02$). Many more good version readers echoed the content of Sentence 1, shown in the first category, than did bad version readers. But many more bad version readers produced responses using the throughout history idea than did good version readers. The propositional analysis of the responses, shown in Table 7, agrees for the most part with the simple categorization. The propositions that appear explicitly in Sentence 1, such as (USE CULTURE METAL) and (MOD METAL DIFFERENT), appear in good version responses much more often than in bad version responses, whereas the (THROUGHOUT P* HISTORY) form is more common in the bad version responses. But the production frequencies differed only marginally between the two versions ($X2(15)=23.716$, $p<.1$).

Table 6

Response Categories for METALS

Version		Category
Good	Bad	
18	6	Different cultures have used different metals for different reasons
9	14	Different cultures value different metals
5	15	Different metals have been used by people throughout history
2	7	Different cultures have valued different metals throughout history.
7	2	Metals have affected the course of human societies
8	6	Metals have many uses and values
3	2	The values of a metal depends on its use
5	5	miscellaneous

Table 7

Propositional Analysis of METALS Responses

Proposition	Production Frequency Good	Bad
(MOD METAL DIFFERENT)	27	19
(MOD CULTURE DIFFERENT)	20	15
(FOR P* PURPOSE)	19	10
(MOD PURPOSE DIFFERENT)	17	9
(POSSESS METAL USE)	16	16
(USE CULTURE METAL)	15	8
(THROUGHOUT P* HISTORY)	11	23
(POSSESS METAL VALUE)	11	10
(USE SOMEONE METAL)	9	6
(MOD USE DIFFERENT)	6	8
(NUMBER-OF PURPOSE MANY)	5	2
(MOD VALUE DIFFERENT)	5	3
(VALUE CULTURE METAL)	8	4
(IN P* CULTURE)	5	4
(NUMBER-OF USE MANY)	5	4

Ratings. Table 8 shows the distribution of importance ratings for each sentence in the two versions along with the modal response for each sentence. The distributions for each sentence were compared using a chi-square test, and the significance of the comparison is shown for each sentence. Sentence 1 is given high central ratings, but notice that Sentence 2 shows no difference in ratings. The immediate implication is that readers can readily distinguish between the general content of Sentence 1 and the specific content of Sentence 2, even when Sentence 2 appears first. The remaining sentences show strong version differences on Sentences 4, 5, and 6, and somewhat on Sentence 9. Examination of the passage (Table 1) suggests that readers in the bad version might entertain a main idea having to do with warfare or cultural conflict, and these sentences are those that either strongly suggest or refute this theme. But in the good version, readers may be protected from this alternate main idea.

Protocols. The protocol summaries are shown in Table 9 together with the modal importance rating. The protocols are represented by symbols that for each sentence and each subject summarize the decision about the status of the sentence, and a possible action involving revision of the reader's main idea. A change in the main idea is shown if it was judged to be a major change; minor revisions were ignored for this table. Included in the table is a key to the symbols.

In the good version, Sentence 1 is accepted by most of the subjects as the main idea, but on the bad version first sentence, Sentence 2, one subject reserves judgement, and the others generalize the sentence. The typical main idea then changes on Sentences 4 and 5, with the typical reported main idea being concerned with warfare or cultural conflict. Then at Sentence 6, good version readers tend to return to the first sentence main idea, and bad version readers also revise their main idea to something similar to Sentence 1. Thereafter, most sentences are either subsumed or irrelevant, and relatively few changes in main idea are reported.

As discussed above, the protocol data is of problematic quality, and the many apparent disagreements with the ratings present some problems. An important problem is that the good version protocol subjects made many revisions in their main ideas compared to the bad version subjects. However, examination of the protocols suggests rather strongly that this is an artifact of the random assignment of subjects to the passage version; At least three of the good version subjects were the most loquacious and active subjects; in particular, subject No. 7 engaged in main ideas that were almost confabulatory in their distance from the actual passage content. Protocol collectors, beware! However, the agreement is quite clear on the irrelevant

Table 8

Importance Ratings for METALS

		Good Version				Bad Version			
Sentence	Sig	C	R	U	Mode	C	R	U	Mode
1.	---	.50	.44	.06	C				
2.	NS	.14	.81	.06	R	.19	.72	.08	R
3.	NS	.14	.81	.06	R	.17	.78	.06	R
4.	**	.00	.14	.86	U	.25	.44	.31	R
5.	**	.11	.81	.08	R	.39	.61	.00	R
6.	**	.08	.61	.31	R	.00	.36	.64	U
7.	NS	.00	.17	.83	U	.03	.11	.86	U
8.	NS	.06	.64	.31	R	.03	.61	.36	R
9.	*	.00	.25	.75	U	.14	.39	.47	U
10.	NS	.19	.69	.11	R	.25	.61	.14	R
11.	NS	.00	.39	.61	U	.06	.33	.61	U
12.	NS	.08	.72	.19	R	.11	.69	.19	R
13.	NS	.00	.03	.97	U	.03	.11	.86	U
14.	NS	.00	.61	.39	R	.06	.44	.50	U

* significant at .05; ** significant at .01; NS: $p > .05$

Table 9

Protocol Summary for METALS

	Good Version						Bad Version					
Sentence	Rating	Subjects					Rating	Subjects				
Number	Mode	1	4	7	9	12	Mode	3	5	6	10	11
1.	C	A	A	C	A	A						
2.	R	S	RC	R	S	S	R	GT	RJ	G	GT	G
3.	R	S	SC	RC	S	S	R	G	G	S	G	S
4.	U	I	R	RC	RC	I	R	R	RC?	R	RC?	RC
5.	R	RC	RC	S	RC	R	R	R	RC	R	RC	RC
6.	R	RC	RC	RC	RC	S	U	RC	RC	S	SC	RC
7.	U	R	I	R	I	I	U	I	I	R	I	I
8.	R	S	RC?	RC	S	R	R	RC?	SC	RC	R	R
9.	U	S	R	S	S	S	U	RC	S	RC	R	R
10.	R	S	S	RC	RC	S	R	S	RC?	S	S	S
11.	U	I	R	S	I	R	U	R	R	S	R	R
12.	R	S	RC?	SC	RC	S	R	R	RC?	S	S	S
13.	U	I	R	S	I	R	U	R	R	S	I?	R
14.	R	RC?	S	R	I	S	U	S	I	S	R	S

Key.
A = accept sentence as statement of main idea
G = generalize this and prior sentences to produce a main idea
GT = generalize to produce a candidate topic for the passage
RJ = reserve judgement about main idea
C = change candidate main idea
C? = state a tentative change
S = judge sentence as subsumed under candidate main idea
R = judge sentence as related to main idea
I = judge sentence as irrelevant to main idea

sentences.

Reading Times. The reading times are shown in Figure 1 for METALS. This shows the "profile" of reading times on each sentence in the passage, for the two versions. Note that the reading times for identical sentences in the two versions are plotted at the same abscissa point. Along the abscissa is an indication of the significance of a t-test for the difference between the version reading times for each sentence. Longer reading times appear for Sentences 4, 5, and 6 in the bad version compared to the good version. Note that this was where revisions were indicated by the ratings and protocol data. Also, Sentence 10 is read longer in the bad version, which is where the warfare theme is finally refuted. Note also the longer reading time on Sentence 2 in the bad version.

The METALS passage shows strong version effects, which have to be attributed to macro-structure processes, since only the first sentence was different. Other passages do not show such effects. A question to ask is how much of the reading time is due to macrostructure processes? One way to see this is to use multiple regression to predict the reading time based on superficial sentence properties. The properties used are WORDS, the number of words in the sentence, FAM, the familiarity ratings, and a dummy variable FIRST, which is equal to 1 on the first sentence in each version and zero otherwise. The analysis was done using the mean reading times for the 88 sentences in both versions of the four passages. The prediction equation is RT = 3.275 - (.333) FAM + (.221) WORDS + (1.773) FIRST. About 84% of the variance for all four passages is accounted for, and all variables contribute significantly at the .01 level. This will be referred to as the WORDS prediction equation. Note that the presence of FIRST in the equation means that generally the first sentence in a passage required substantially longer to read than can be predicted just on the basis of its length or familiarity. The presence of FAM means that more familiar sentences took less time to read, with length taken into account.

The predicted and observed times for METALS are shown in Figures 2A and 2B. Sentences 2, and 4 in the bad version, and 5 and 10 are being read for different amounts of time than would be expected based on these superficial properties. Sentences 2, 4, and 10, where revisions seem to be required, are read longer. Sentence 5 appears to be a special case; it is very long, but contains very little "new" information (see Kieras, 1978, 1981), and so is read for less time than would be expected.

Summary. So, in METALS, in the bad version, readers consider revisions frequently, but their candidate main ideas produced during reading apparently converge through the course of the passage, to the same main idea presented

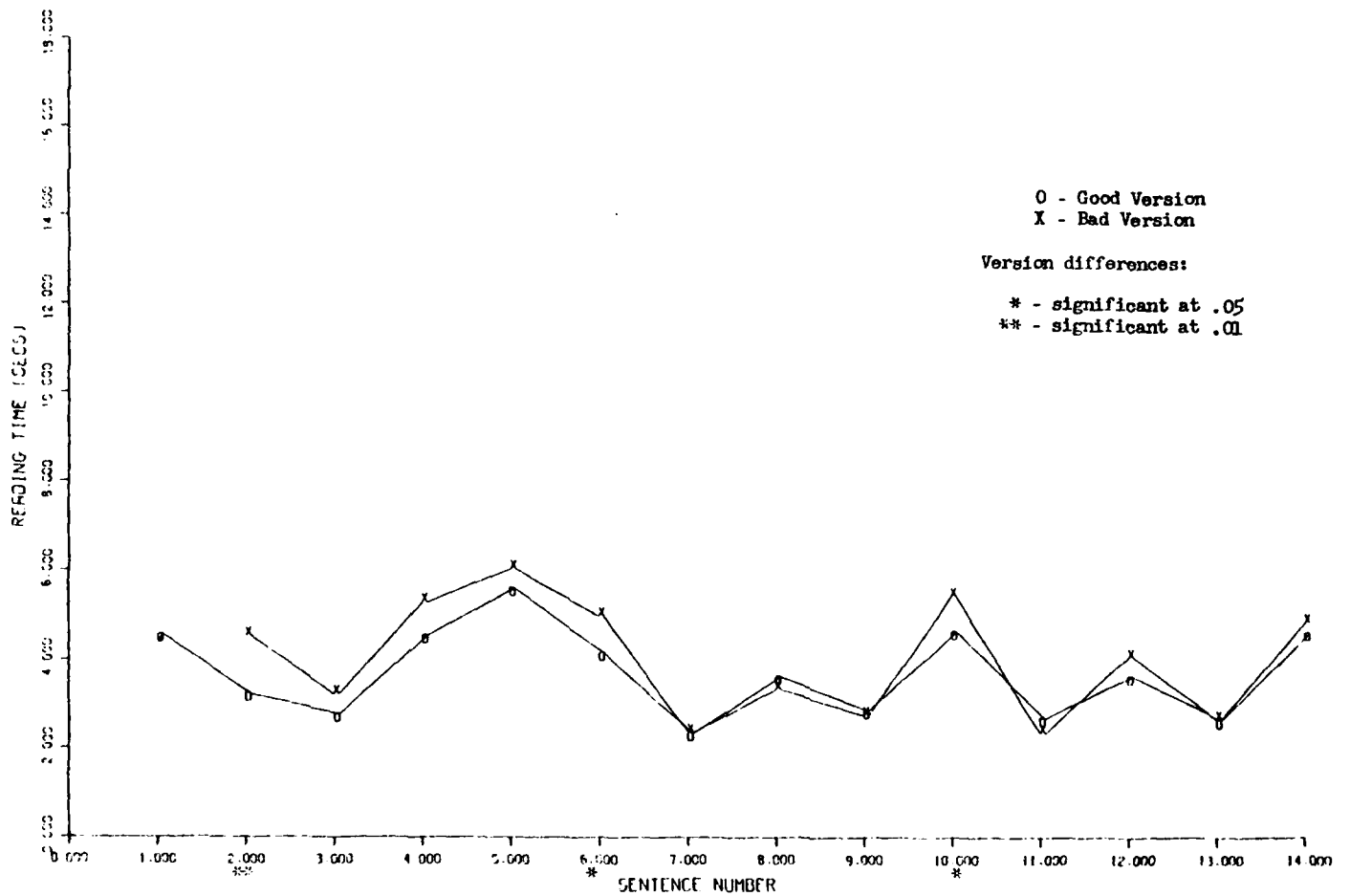


FIGURE 1. READING TIMES FOR METALS PASSAGE

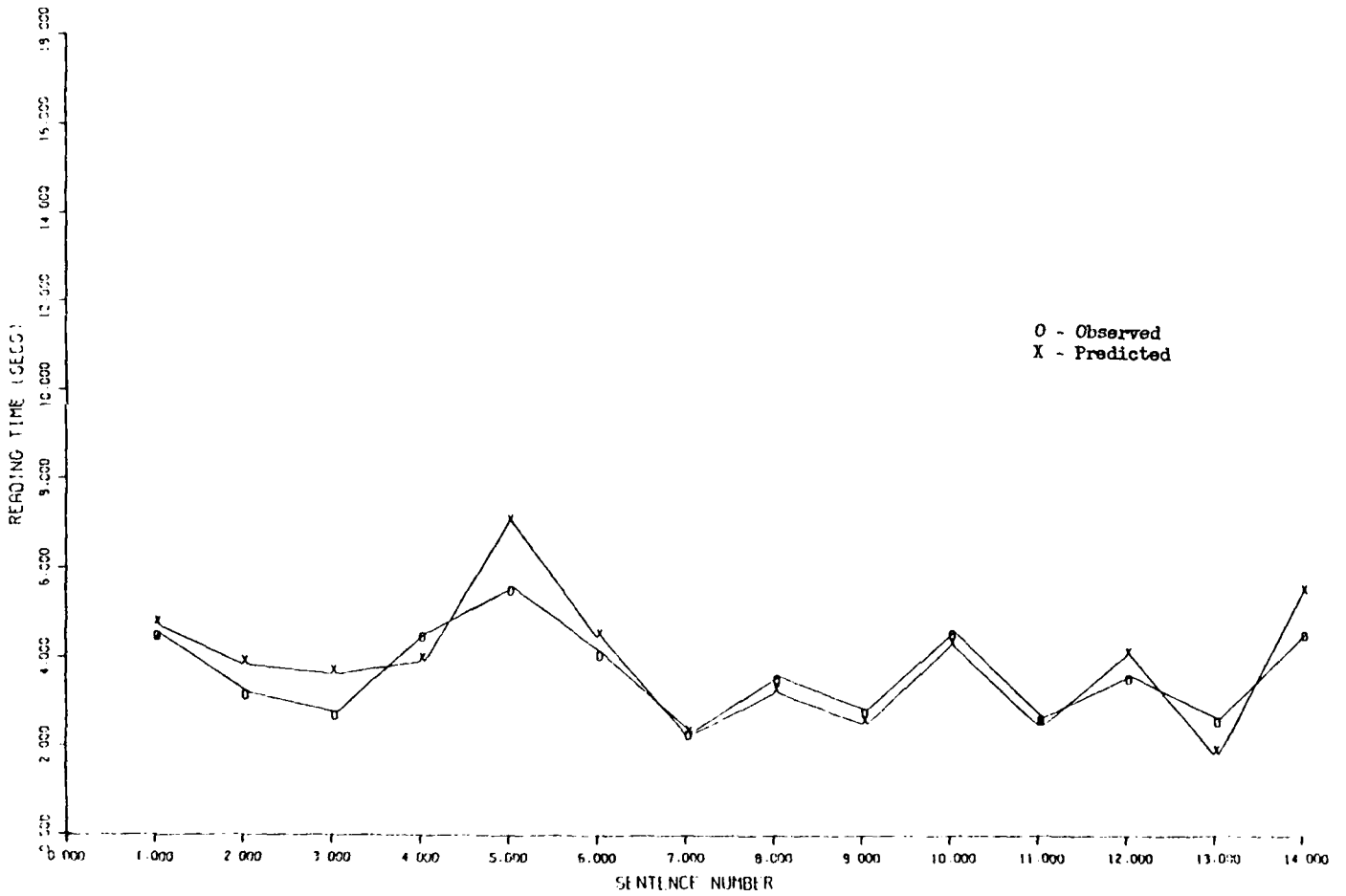


FIGURE 2A. PREDICTED (WORDS) VS OBSERVED TIMES FOR METALS, GOOD VERSION

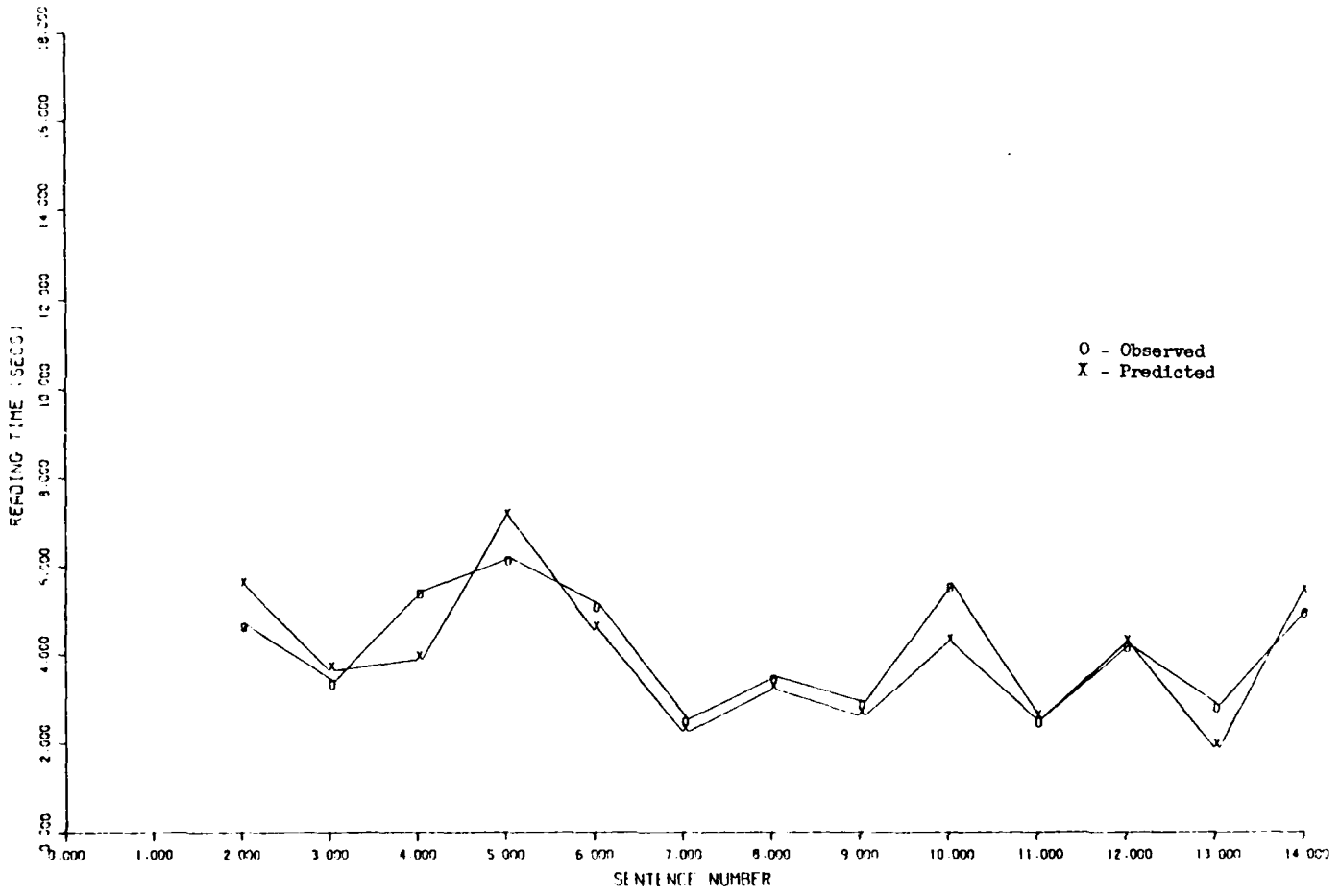


FIGURE 28. PREDICTED (WORDS) VS OBSERVED TIMES FOR METALS. BAD VERSION

in the good version. The final responses are thus very similar in content; the only important difference is in the case of "throughout history" in the bad version responses. The good version first sentence is recognized as a good candidate main idea, and is echoed in many of the responses, while the bad version first sentence is recognized as not being a good main idea. Where revisions are often involved, we see longer reading times.

The TIMEKEEPING passage

Responses. The response categorization is shown in Table 10, in which the difference in version distributions was significant ($X2(9)=24.927$, $p<.01$). Note again how many good version readers simply echoed the first sentence. A large number of bad version readers produced a good generalization such as those in the second, third, and fifth categories, but there were also several responses focused on specific items, such as the hydrogen maser. The propositional analysis is shown in Table 11. Again the good version readers used propositions explicitly contained in Sentence 1, while bad version readers used a more diffuse set of propositions, being recognizable portions of the other responses shown in Table 10. The production frequencies of the propositions were significantly different for the two versions ($X2(16)=59.935$, $p<.001$).

Ratings. The importance ratings in Table 12 show that Sentence 1 is again given high central ratings, but Sentence 2 is not in the bad version. Sentence 3, a detail about quartz-crystal watches is more important in the bad version than in the good, suggesting that bad version readers may have taken this item as the passage topic. Sentence 5 is heavily judged irrelevant, but more so in the good compared to the bad version. The remaining sentences are all judged important, and show no version effects.

Protocols. In the protocols (Table 13), the first sentence is again accepted outright in the good version. In the bad version, several subjects generalized Sentence 2, arriving at ideas such as how Man measures time.Sentence 5, the large sentence about relativity, produced few revisions in the good version, caused many revisions in the bad version, which were then abandoned on the next sentence.

Reading times. The reading times (Figure 3) for this passage showed no version differences, except for the hint that the bad version Sentence 2 is read longer than in the good version. This lack of effect can be explained by the fact that in the think-aloud protocols, many readers made a good guess at the intended main idea very early in the passage. If so, then the bad version reader will be essentially in the same state as if the intended main idea had been explicitly presented, and hence no version effects

Table 10

Response Categorization for TIMEKEEPING

Version		Category
Good	Bad	
22	4	Modern timekeeping devices are extremely accurate
7	15	Different timekeeping devices have different degrees of accuracy
9	10	Clocks can be very accurate
4	3	Clocks are more accurate today than in the past
2	7	Modern technology has improved the accuracy of clocks
1	5	The hydrogen maser clock is the most accurate clock
1	2	What a clock is used for depends on its accuracy
1	2	Clocks can be used to support the tneory of relativity
3	0	This passage was about the accuracy of various timepieces
7	9	miscellaneous

Table 11

Propositional Analysis of TIMEKEEPING Responses

Proposition	Production Frequency Good	Bad
(MOD P* EXTREMELY)	24	10
(MOD TIMEPIECE ACCURATE)	25	8
(TIME TIMEPIECE TODAY)	23	3
(POSSESS TIMEPIECE ACCURACY)	12	18
(WITH P* ACCURACY)	10	8
(MEASURE TIMEPIECE TIME)	9	5
(DEGREE-OF ACCURACY EXTREME)	7	4
(TIME P* TODAY)	5	4
(ABLE TIMEPIECE P*)	5	7
(DEGREE-OF ACCURACY DIFFERENT)	4	10
(USE SOMEONE TIMEPIECE)	4	8
(MOD TIMEPIECE DIFFERENT)	3	7
(POSSESS TIMEPIECE TYPE)	2	6
(MOD TYPE DIFFERENT)	1	6
(ABLE SOMEONE P*)	2	5
(MORE-ACCURATE-THAN TIMEPIECE1 TIMEPIECE2)	2	5

Table 12

Importance Ratings for TIMEKEEPING

		Good Version				Bad Version			
Sentence	Sig	C	R	U	Mode	C	R	U	Mode
1.	---	.72	.28	.00	C				
2.	NS	.39	.56	.06	R	.39	.58	.03	R
3.	*	.06	.78	.17	R	.25	.72	.03	R
4.	NS	.42	.56	.03	R	.22	.64	.14	R
5.	*	.03	.11	.86	U	.06	.39	.56	U
6.	NS	.11	.75	.14	R	.31	.61	.08	R
7.	NS	.31	.69	.00	R	.22	.69	.08	R
8.	NS	.11	.44	.44	R,U	.03	.50	.47	R

* significant at .05; ** significant at .01; NS: $p > .05$

Table 13

Protocol Summary for Timekeeping Passage

	Good Version						Bad Version					
Sentence	Rating	Subjects					Rating	Subjects				
Number	Mode	1	5	6	10	11	Mode	3	4	7	9	12
1.	C	AC?	A	A	C	A						
2.	R	S	S	S	S	R	R	G	GT	G	RJ	G
3.	R	I	R	R	R	R	R	R	R	R	R	R
4.	R	S	S	S	S	S	R	S	SG	S	SG	RC
5.	U	R	RC?	R	R	I	U	RC	R	RC	I	R
6.	R	R	RC	R	RC	R	R	RC?	RC?	RC	RC	R
7.	R	S	SC	S	SC	S	R	R	SC	R	S	SC
8.	R,U	I	S	R	I	I	R	R	R	RC	R	R

Key.
A = accept sentence as statement of main idea
G = generalize this and prior sentences to produce a main idea
GT = generalize to produce a candidate topic for the passage
RJ = reserve judgement about main idea
C = change candidate main idea
C? = state a tentative change
S = judge sentence as subsumed under candidate main idea
R = judge sentence as related to main idea
I = judge sentence as irrelevant to main idea

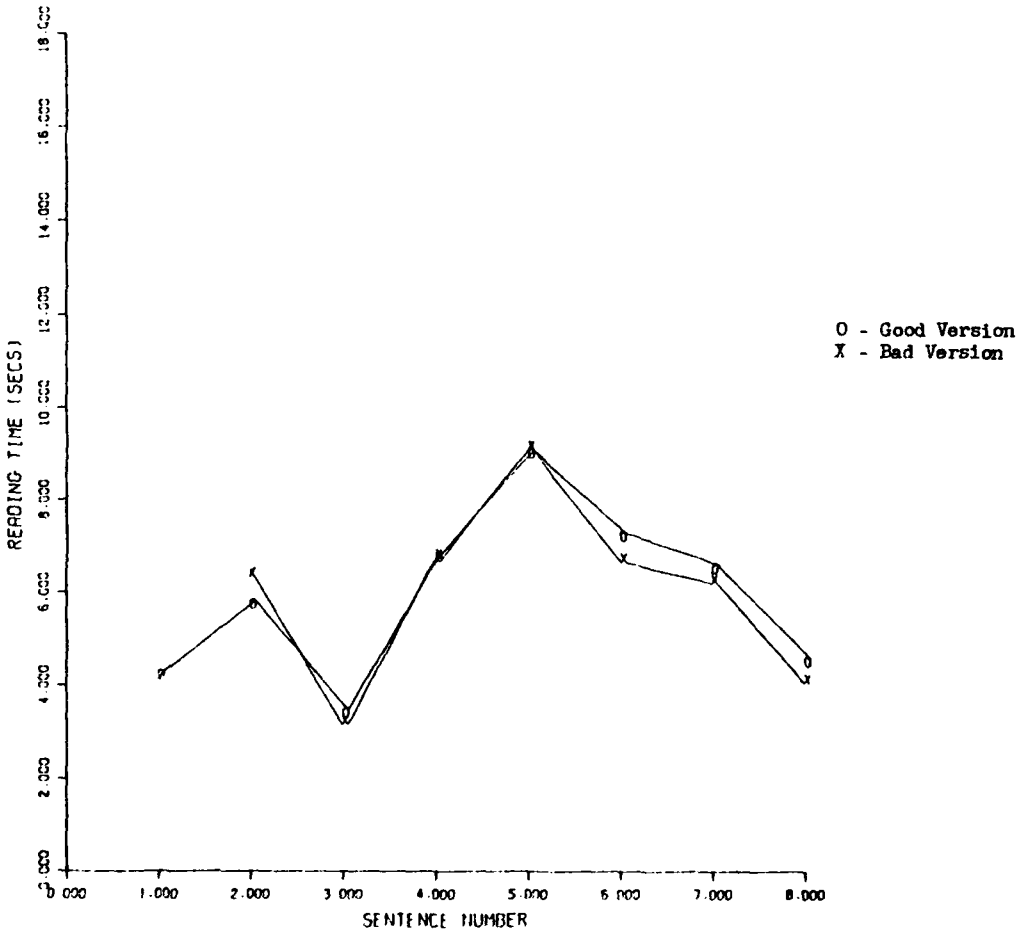


FIGURE 3. READING TIMES FOR TIMEKEEPING PASSAGE

appear. Using the WORDS prediction equation, (Figures 4A and 4B), we see that Sentence 5 is read much longer than would be expected, although it was judged irrelevant, which is consistent with the extensive consideration given to this sentence by the protocol subjects.

Summary. So in the TIMEKEEPING passage the explicit main idea plays a guiding role, but it seems to be quickly inferred if absent. When a large, but irrelevant, sentence appears, such as Sentence 5, it is taken very seriously, and revisions are considered, but not necessarily made.

The INSTRUMENTS passage

Responses. Subjects tended to complain about this passage, saying that it was the hardest of the set, perhaps as a result of the very complex main idea in Sentence 1. The responses shown in Table 14 differ in distribution between versions ($X2(6)=14.289$, $p<.05$). Basically, good version readers reproduce one of two subsets of the content of Sentence 1, whereas bad version readers had a strong tendency to view the passage as about the three specific instruments. The propositional analysis is shown in Table 15. The production frequencies for this passage are generally very low compared to the other passages, especially in the bad version. This indicates a relatively high degree of inconsistency and idiosyncrasy in the responses. However, note that most of the propositions shown (which meet the minimum frequency criterion of 5) are from Sentence 1, and they are produced much less often in the bad version ($X2(9)=32.710$, $p<.01$). The conclusion is that bad version readers are almost unable to agree on the specific content of their main ideas, but show some agreement on the general content of their responses that shows up in the simple categorization analysis.

Importance Ratings. The importance ratings, in Table 16, again show that the explicit main idea presented in Sentence 1 is given high Central ratings. But Sentence 2 in the bad version is also considered to be fairly central. Examination of the passage (Table 3) suggests that Sentence 2 is in fact a good topic sentence about the clavichord, and there is then a tendency to down-play Sentences 6 and 7 about the harpsichord. Perhaps the most striking feature about the ratings for this passage is that all of the sentences are judged highly important; this may account for the relative difficulty of this passage, since all of the information would have to be processed.

Protocols. In the protocols summarized in Table 17, Sentence 1 is accepted, as in the other passages, as stating a main idea. In the bad version, Sentence 2 is accepted or generalized to a passage topic, corresponding to its high central rating. The common hypothesized topic at this point

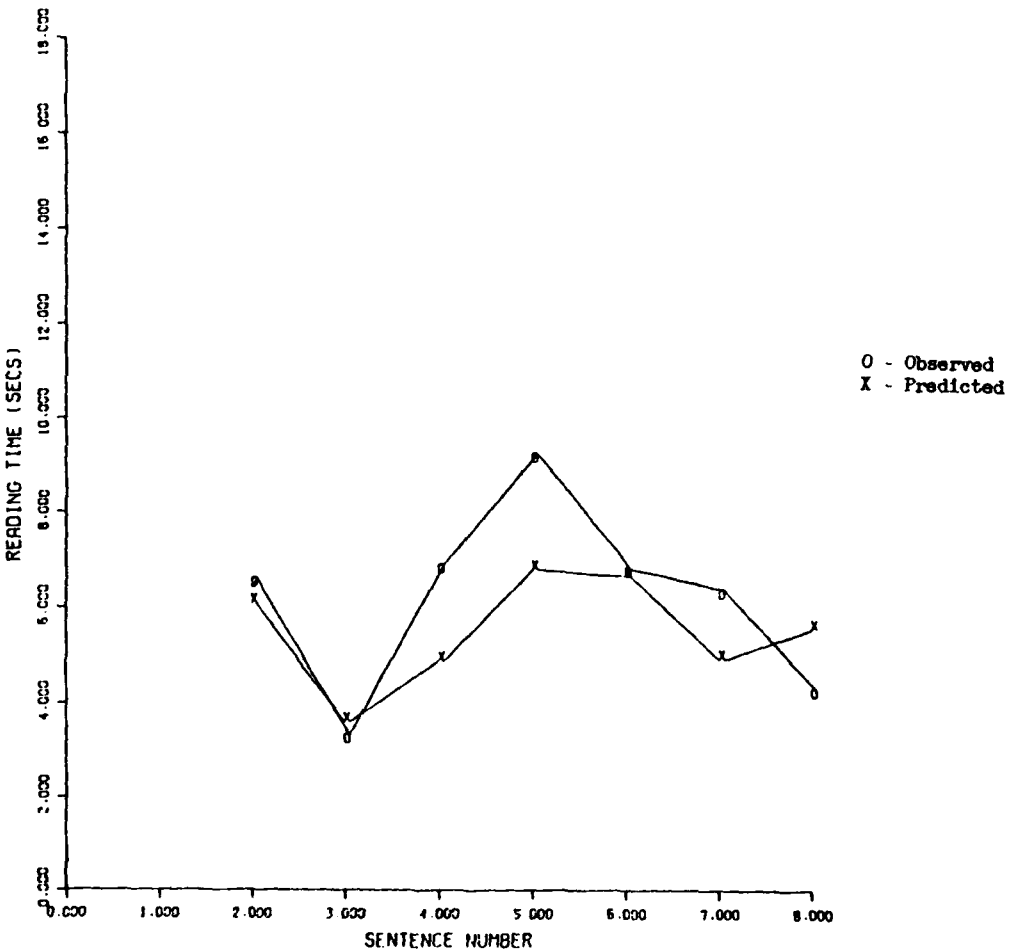


FIGURE 4B. PREDICTED (WORDS) VS OBSERVED FOR TIMEKEEPING. BAD VERSION

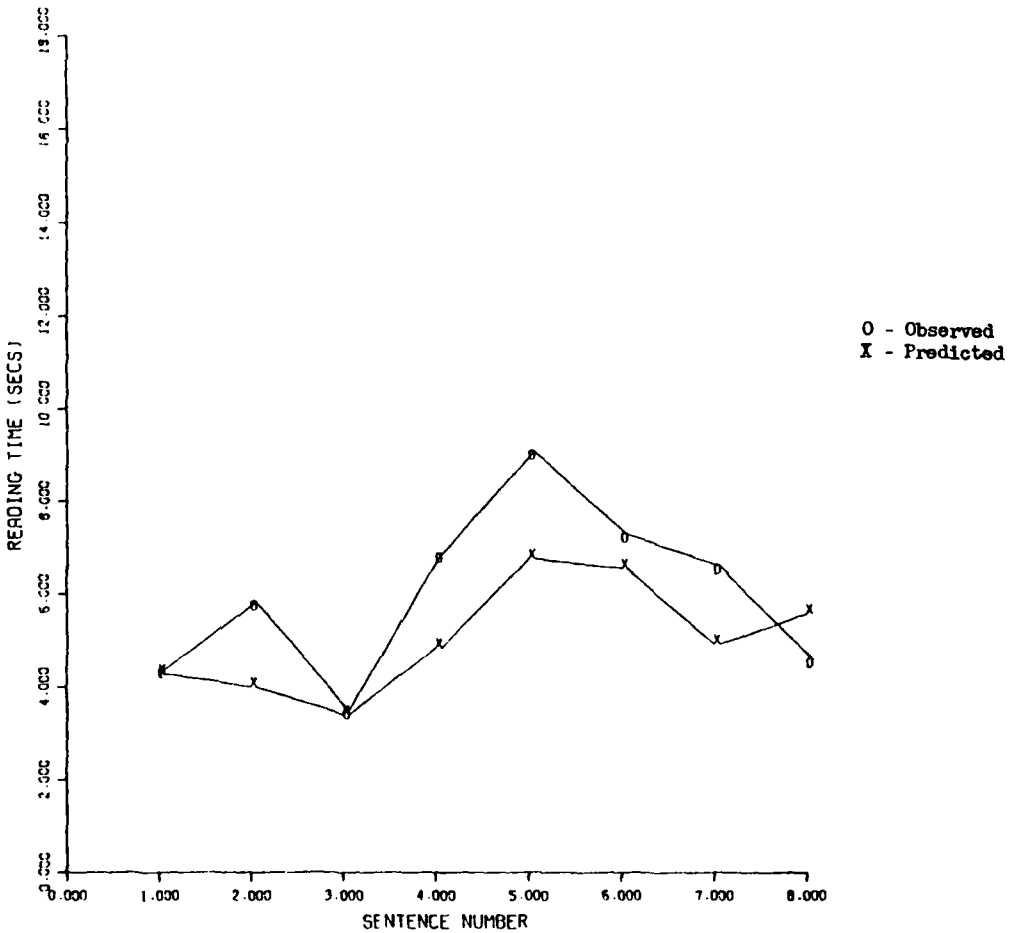


FIGURE 4A. PREDICTED (WORDS) VS OBSERVED FOR TIMEKEEPING. GOOD VERSION

Table 14

Response Categorization for INSTRUMENTS

Version		Category
Good	Bad	
18	10	Different keyboard instruments permit different degrees of control over sound quality
11	3	Differences in the sounds produced by keyboard instruments are due to differences in their mechanisms
4	10	The clavichord, harpsichord, and piano have similar mechanisms
5	13	The piano is superior to the clavichord and harpsichord
4	7	The clavichord, harpsichord, and piano are different
1	2	The clavichord and harpsichord were forerunners of the piano
14	12	miscellaneous

Table 15

Propositional Analysis of INSTRUMENTS Responses

Proposition	Production Frequency Good	Bad
(MOD INSTRUMENT KEYBOARD)	36	16
(ABLE PERFORMER P*)	7	3
(POSSESS INSTRUMENT MECHANISM)	9	3
(ABLE SOMEONE P*)	5	3
(MOD SOUND DIFFERENT)	7	3
(MOD MECHANISM DIFFERENT)	7	3
(PRODUCE INSTRUMENT SOUND)	5	0
(ON P* INSTRUMENT)	5	1
(MOD INSTRUMENT MUSICAL)	2	5

Table 16

Importance Ratings for INSTRUMENTS

		Good Version				Bad Version			
Sentence	Sig	C	R	U	Mode	C	R	U	Mode
1.	---	.67	.31	.03	C				
2.	**	.08	.61	.31	R	.53	.42	.06	C
3.	NS	.11	.75	.14	R	.08	.75	.17	R
4.	NS	.14	.83	.03	R	.08	.92	.00	R
5.	NS	.22	.72	.06	R	.14	.78	.08	R
6.	*	.08	.83	.08	R	.17	.53	.31	R
7.	*	.17	.81	.03	R	.06	.69	.25	R
8.	NS	.17	.72	.11	R	.19	.56	.25	R
9.	NS	.06	.89	.06	R	.11	.81	.08	R
10.	NS	.22	.72	.06	R	.19	.72	.08	R
11.	NS	.19	.44	.36	R	.25	.28	.47	U

* significant at .05; ** significant at .01; NS: $\underline{p} > .05$

Table 17

Protocol Summary for Instruments Passage

Sentence Number	Good Version Rating Mode	Good Version Subjects 3	5	6	9	11	Bad Version Rating Mode	Bad Version Subjects 1	4	7	10	12
1.	C	A	A	AC	A	A						
2.	R	I	R	R	I	R	C	G	GT	G	GT	AT
3.	R	S	S	R	S	S	R	RC	R	RC	RC	S
4.	R	RC?	S	S	S	S	R	S	RC?	RC?	R	R
5.	R	R	S	S	S	S	R	S	R	RC	S	S
6.	R	S	S	S	S	S	R	SC	SC	SC	SC	SC
7.	R	S	S	R	S	R	R	S	RC	RC	R	SC
8.	R	S	S	S	S	S	R	S	SC	S	S	SC
9.	R	R	S	S	S	R	R	S	S	R	S	S
10.	R	R	S	R	R	S	R	S	R	S	S	S
11.	R	I	RC?	R	I	R	U	R	S	S	S	S

Key.
A = accept sentence as statement of main idea
G = generalize this and prior sentences to produce a main idea
GT = generalize to produce a candidate topic for the passage
RJ = reserve judgement about main idea
C = change candidate main idea
C? = state a tentative change
S = judge sentence as subsumed under candidate main idea
R = judge sentence as related to main idea
I = judge sentence as irrelevant to main idea

is the clavichord. In the good version, very few revisions occur, perhaps because the subjects could not engage in as much speculative inference on this unfamiliar passage compared to the more familiar ones. Sentences 6 and 7 produce no changes in the good version because they are related to the main idea, but in the bad version at Sentence 6 everyone revises their main idea. The people with the clavichord topic abandon their hypothesis and adopt a more general one, such as keyboard instruments.

Reading times. The reading times are not reported because they show no interesting version differences, and the regression analysis is not informative.

Summary. INSTRUMENTS behaved substantially like the other passages, but it is substantially harder than the other passages, it is one of the least familiar, and has a very high proportion of content that is important. Relatively few readers acquired the intended main idea, but more did so in the good version. In the bad version, the first sentence was treated as a good topic sentence, but readers still had to revise their main idea.

The CARS passage

Responses. As shown in Table 18, the two versions produced a similar distribution of responses ($X2(7)=10.012$, $p<.25$). The good version readers echoed most of the main idea sentence, and bad version readers did also, but to a somewhat lesser extent. The propositional analysis of the responses (Table 19) is entirely consistent; very few notable differences appeared between the versions ($X2(17)=18.462$, $p<.5$). Apparently, subjects were able to infer the intended main idea from the bad version as readily as from the good.

Ratings. Again Sentence 1 is highly central, but Sentence 2 in the bad version is also recognized as a good topic sentence and given high central ratings. Sentence 4, which mentions features of imported luxury cars, is more important in the bad version, which suggests that many readers think the passage topic is luxury cars. But Sentences 5, 9 and 12 mention instances that can not be subsumed under the luxury car topic. In the bad version there is a tendency to downrate these sentences compared to the good version. However, Sentence 10, and to some extent, Sentence 13, are more important in the bad version. These sentences deal with the imported-domestic issue, suggesting that bad version readers consider it important.

Protocols. In the protocols (Table 21), the good version readers accept Sentence 1 and make few changes thereafter. In the bad version, the initial Sentence 2 is also accepted, with luxury cars as the topic, but at

Table 18

Response Categorization for CARS

Version		Category
Good	Bad	
24	17	Different people prefer different cars
11	14	Automobile preference is a function of automobile purpose
6	8	Different cars serve different purposes
6	3	Lifestyle determines automobile preference
1	6	There are many types of cars
4	1	Automobiles are preferred for their features
1	4	People prefer cars for many reasons
4	4	Miscellaneous

Table 19

Propositional Analysis of CARS Responses

Proposition	Production Frequency Good	Bad
(IN-ORDER-TO P* P*)	14	18
(BUY PEOPLE CAR)	21	22
(POSSESS CAR TYPE)	19	19
(MOD TYPE DIFFERENT)	13	8
(POSSESS PEOPLE NEED)	13	21
(PREFER PEOPLE CAR)	12	7
(POSSESS PEOPLE DESIRE)	12	13
(MOD PEOPLE DIFFERENT)	10	10
(SUIT CAR NEED)	8	12
(SUIT CAR DESIRE)	7	9
(EXIST CAR)	6	9
(SUIT CAR PEOPLE)	5	6
(MOD CAR DIFFERENT)	5	15
(MOD REASON DIFFERENT)	5	5
(POSSESS CAR FEATURE)	5	1
(MAKE SOMEONE CAR)	4	7
(POSSESS PEOPLE LIFESTYLE)	4	5

Table 20

Importance Ratings for CARS

		Good Version				Bad Version			
Sentence	Sig	C	R	U	Mode	C	R	U	Mode
1.	---	.64	.33	.03	C				
2.	**	.11	.81	.08	R	.50	.44	.06	C
3.	NS	.08	.75	.17	R	.06	.69	.25	R
4.	*	.06	.61	.33	R	.28	.58	.14	R
5.	**	.28	.67	.06	R	.08	.44	.47	U
6.	NS	.03	.22	.75	U	.06	.28	.67	U
7.	NS	.28	.69	.03	R	.25	.64	.11	R
8.	NS	.17	.50	.33	R	.14	.64	.22	R
9.	*	.11	.78	.11	R	.11	.53	.36	R
10.	**	.06	.31	.64	U	.08	.64	.28	R
11.	NS	.00	.33	.67	U	.08	.42	.50	U
12.	*	.28	.67	.06	R	.19	.47	.33	R
13.	NS	.00	.36	.64	U	.06	.53	.42	R

* significant at .05; ** significant at .01; NS: $p > .05$

Table 21

Protocol Summary for Cars Passage

Sentence	Good Version						Bad Version					
	Rating	Subjects					Rating	Subjects				
Number	Mode	3	4	7	10	12	Mode	1	5	6	9	11
1.	C	A	A	C	A	A						
2.	R	S	R	R	S	I	C	A	A	A	A	A
3.	R	S	S	S	S	C	R	R	S	S	S	RC
4.	R	S	R	R	S	S	R	R	R	S	S	R
5.	R	S	R	S	S	S	U	RC	S	SC	IC?	SC
6.	U	I	I	S	S	I	U	S	I	I	RC?	I
7.	R	S	R	S	S	S	R	S	SC	S	S	S
8.	R	S	R	S	S	S	R	S	S	S	S	R
9.	R	S	R	S	S	S	R	S	S	S	S	R
10.	U	I	I	R	R	I	R	I	I	R	I	R
11.	U	I	R	R	R	R	U	R	R	R	RC	R
12.	R	S	S	S	S	S	R	S	S	S	R	S
13.	U	R	R	R	R	R	R	R	R	R	RC?	R

Key.
A = accept sentence as statement of main idea
G = generalize this and prior sentences to produce a main idea
GT = generalize to produce a candidate topic for the passage
RJ = reserve judgement about main idea
C = change candidate main idea
C? = state a tentative change
S = judge sentence as subsumed under candidate main idea
R = judge sentence as related to main idea
I = judge sentence as irrelevant to main idea

Sentence 5, about station wagons, readers change their main idea. An interesting exception is one subject who subsumed this sentence under the luxury car topic, saying that station wagons were in fact luxury cars. This subject undertook a complete revision when the pickup sentence apeared.

Reading times. Again the reading times show no version effects, and no interesting deviations from the words predictions.

Summary. The CARS passage was the most familiar in content. Perhaps as a result, readers were able to acquire the intended main idea in both versions equally well, with no differences in reading times, and few differences otherwise. Like INSTRUMENTS, the bad version first sentence was adopted as a good main idea, and then later rejected.

Reader Strategies

The Subsuming Strategy. The overall strategy that most readers seem to use can now be stated. The first sentence is tested to see if it appears to express a reasonable main idea. This test uses only the superficial characteristics of the sentence, such as whether general concepts are referred to, and so can be performed immediately and without any prior context. If the first sentence is general, it is adopted as the candidate main idea, and the reader attempts to "fit," or subsume, each succeeding sentence into this main idea. If this attempt begins to fail at some point in the passage, revisions in the candidate main idea will be considered, and possibly carried out. In this strategy, the key operation is that of subsuming each sentence under the main idea, so it will be called the subsuming strategy.

Basically, the two passage versions are treated differently by the subsuming strategy in the following way: In the good version, revisions are usually not necessary, since the main idea stated in the first sentence actually subsumes most of the remaining sentences. But in the bad version, several revisions might be necessary. Since the passages in the bad version are generated from a generalization (which is stated in the good version) the revisions tend to eventually arrive at this generalization.

Defective strategies

In all of the studies done by Kieras (Note 1, Note 4, 1980, 1981) of how people abstract thematic content from passages, many instances of very poor-quality responses have been observed. These could be either (a) the result of awkward verbal expression skills on the part of subjects, or (b) subjects making very poor choices of the content to include in their response. If it is a matter of poor verbal expression skills, the problem of poor responses is the same one as why many students can not write. But the problem of poor choice of content would seem to reflect problems in basic reading comprehension skill, and should thus be due to defective strategies.

In order to study poor readers, the first step is to define them. The definition used here was based on the fact that generally large numbers of the readers could produce the intended main idea of a passage even in the bad version. Hence the extent to which readers reproduced the intended main idea in their response was the initial distinction between good and bad readers. Each response was classified as being good if it reproduced most of the propositions from the intended main idea sentence, fair if it reproduced only the main proposition of the intended main idea, and poor if it failed to contain the main propostion. Some examples of good and poor responses are shown in Table 22. Subjects were then designated as good, fair or poor readers, based on the response classification, for each passage.

Notice that there are actually two different types of poor readers. In the good version, these were readers who missed an explicitly stated main point. In the bad version, poor readers failed to draw the same inference as good readers did. Thus, as would be expected, the classification produced more poor subjects in the bad versions, but only in the least familiar passages, TIMEKEEPING and INSTRUMENTS. In the METALS and CARS passages, there was no difference in the proportions of good, fair, and poor readers.

The question was whether there were any differences on any other measures between good and poor readers. The initial results were very discouraging. The mean reading times were almost identical for good and poor readers. The distributions of mean reading times also showed no difference. The familiarity ratings showed no good-poor difference either, which would be expected, perhaps, from the conclusion (see below) that only modest amounts of knowledge are needed to successfully process the passages in the main idea task. Another attempt consisted of purifying the groups by including only subjects who were either consistently good or consistently poor, defined as being classified the same way on three of the four passages. The mean reading times, the profile of reading times, and the mean and single-sentence familiarity ratings were almost

Table 22

Examples of Good, Poor, and Focus responses to METALS

GOOD:

Different cultures used different metals for a variety of reasons.

Different metals are valued for varying reasons.

POOR:

Man has a multitude of uses for metals.

Men regard the importance of metals according to their uses.

FOCUS:

The Incas loved gold; whereas the Spaniards did also and conquered them.

Materials used in ancient wars are now expensive and scarce.

---

identical for even these two groups. A next attempt consisted of classifying each sentence as being either important or unimportant based on the importance rating data, and then looking for good-poor reader differences in the reading times. No difference was obtained.

The next step was to focus on the importance ratings themselves, based on the idea that good and poor subjects might not differ in how long they read each sentence, but rather in the importance they attach to individual sentences. For example, in the good version of the METALS passage, poor subjects tended to rate Sentence 6, about the Incas using gold, as less important than good subjects did. This is the first sentence that disconfirms the warfare theme, and so should be judged as directly related to the intended main idea. Hence, perhaps poor subjects do not weight new evidence that they encounter in the passage as efficiently as good readers. Likewise, in the bad version, poor readers make more central judgements on Sentence 4, about the Hellenes invading Greece, and thus are not using the common weapons generalization that many subjects inferred in the first two sentences.

The appearance of differences in importance ratings led to the consideration of a more specialized form of poor subject, which could be related more exactly to the importance ratings. These subjects are termed "focusers" because they focus on a specific fact in the passage, and so produce a very specific response, rather than a generalization. Some examples of such focus responses are shown in Table 22.

The focusers do not generalize the passage content, but rather insist on summarizing the passage in terms of a specific item. It is unlikely that they are simply sloughing the task, because almost all of the focusers show as much variety in importance ratings as ordinary subjects. Hence, they must be seriously working on the passage, but follow a rather different strategy for abstracting the main idea. Some examples of how the responses can be tied to differences in importance ratings for the good version of the TIMEKEEPING passage, are shown in Table 23, which shows the importance ratings given by a group of focusers and by the good subjects. The focusers rate the first sentence as less important than the good subjects do, but judge Sentence 5, Sentence 6, and Sentence 8 as more important. Correspondingly, several focus responses are about hydrogen maser clocks, and how clocks are used to test the theory of relativity.

Thus, one feature of focusers, compared to most readers, seems to be a different set of rules for using the first sentence. For such readers, the intended main idea sentence in the good version is not apparently recognized as such, since it is rated relatively low in importance, and in

Table 23

Importance ratings for Good and Focus subjects on TIMEKEEPING

Good version

Sent. No.	Good Subjects C	R	U	Focus Subjects C	R	U	Sig.
1.	.74	.26	.00	.40	.60	.00	NS
2.	.44	.48	.07	.40	.60	.00	NS
3.	.07	.82	.11	.00	.80	.20	NS
4.	.52	.44	.04	.20	.80	.00	NS
5.	.00	.11	.89	.20	.00	.80	*
6.	.11	.74	.15	.20	.80	.00	NS
7.	.41	.59	.00	.00	1.00	.00	NS
8.	.04	.48	.48	.60	.40	.00	**

Bad version

Sent. No.	Good Subjects C	R	U	Focus Subjects C	R	U	Sig.
2.	.33	.67	.00	.25	.67	.08	NS
3.	.42	.58	.00	.08	.83	.08	NS
4.	.17	.58	.25	.25	.58	.17	NS
5.	.00	.42	.58	.17	.33	.50	NS
6.	.17	.75	.08	.67	.25	.08	*
7.	.25	.75	.00	.08	.67	.25	NS
8.	.00	.67	.33	.08	.42	.50	NS

* significant at .05; ** significant at .01; NS: $p > .05$

the bad version, Sentence 2 is often over-rated in importance. This suggests that focusers are less sensitive to the generalization content of sentences, especially the passage's initial topic sentence. The second feature is that apparently they do not use the subsuming strategy, because often sentences closely related to the intended main idea are down-rated by focusers compared to good subjects, and specific item sentences are highly rated.

A SIMULATION MODEL OF THE SUBSUMING STRATEGY

The simulation is essentially a production-system version of van Dijk's macrostructure building rules (van Dijk, 1977a, b; 1980). The simulation starts with a propositional representation of the content of each sentence, based on Kintsch (1974), and processes one sentence at a time, and attempts to extract a generalization. The input to the model is the list of propositions in the passage, segmented by sentence.

The model consists of several sets of production rules arranged hierarchially. The top level is a set of control productions that cause the processing to proceed one sentence at a time. This top level invokes additional sets of production rules to carry out the processing. One set handles the first-sentence special case, another controls the processing on each sentence thereafter. Other sets perform the subsumption testing and generation of a new generalization and classifying the sentence propositions. Finally, another rule set performs the crude inferential processing required before many of the sentences can be tested for subsumption. The model is run by a specialized production system interpreter written in LISP. Further details of the model implementation will not be described here; copies of the LISP source listings for the model and the interpreter are available from the first author.

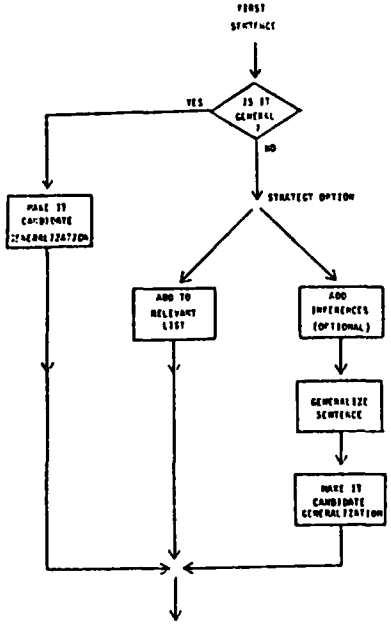
The model assumes several memory systems, each consisting of a list of propositions. The long-term memory (LTM) consists of a list of propositions stating general knowledge. This list is prepared separately for each passage, and so the model has only one passage's worth of general knowledge at a time. The LTM propositions consist mostly of ISA relationships defining set membership and various IMPLY propositions which are used by the inference production rules. The working memory (WM) contains all of the sentence propositions that the model has seen while processing the passage, and also the propositions created while generating inferences and generalizations. The WM is subdivided into several lists, one for the candidate main idea, and other lists for the previously classified input. The content of these lists indicates which propositions were subsumed, which were found related to the main idea, and which were irrelevant. Finally, short-term memory could be

said to be represented as the many temporary lists that are constructed in the course of processing for purposes such as keeping track of intermediate results while generating a generalization from a list of propositions.

The subsuming strategy is implemented by a straightforward set of production rules, summarized in the flowchart (Figure 6). The first sentence is accepted as general if the main proposition of the sentence contains general terms, and is used as the first candidate generalization main idea. If the first sentence is not general, the system either waits for the next sentence, or generalizes the first sentence by replacing the main proposition with one in which general terms replace the specific ones. The system classifies each succeeding sentence into one of three categories. The sentence might be subsumed, in that it contains a proposition that is an instance of the current candidate main idea generalization, or it may be related to the main idea by sharing terms with propositions that are already subsumed or related, or it is irrelevant, neither subsumed nor related. After classifying the sentence, the system then decides whether enough of the passage content is still subsumed. If so, it goes on to the next sentence. If not, it generates a new candidate generalization from the content of all propositions processed thus far. It then reclassifies the previous passage content, and continues. At the end of the passage, the model reports its current candidate generalization as its main idea for the passage.

In developing the model, the first goal was to enable the model to produce a main idea proposition that at least roughly resembled the main propositions appearing most often in the subjects' responses. Note that the many auxiliary modifying propositions that the subjects use in their responses are not generated by the model; it develops a single proposition that represents its final candidate main idea. The question in evaluating the model's realism is then not the quality of the final main idea, but the similarity of the sentence-by-sentence processing to the subjects' ratings, protocols, and reading times.

Once the simulation could generate reasonable main idea propositions, it became clear that the central problem in making the model realistic was the criterion used to decide whether enough was subsumed. There are many possibilities, but the approach reported here was based on making the decision on the basis of the relative number of propositions in the sentence and in the subsumed, related, and irrelevant lists, along with the classification of the sentence. For example, a useful overall rule is that if a very large irrelevant sentence appears, a revision should be attempted. It was quickly found that the most promising criteria are "dynamic" in the sense that the nature of the first sentence determines the specific criterion used in the rest of the



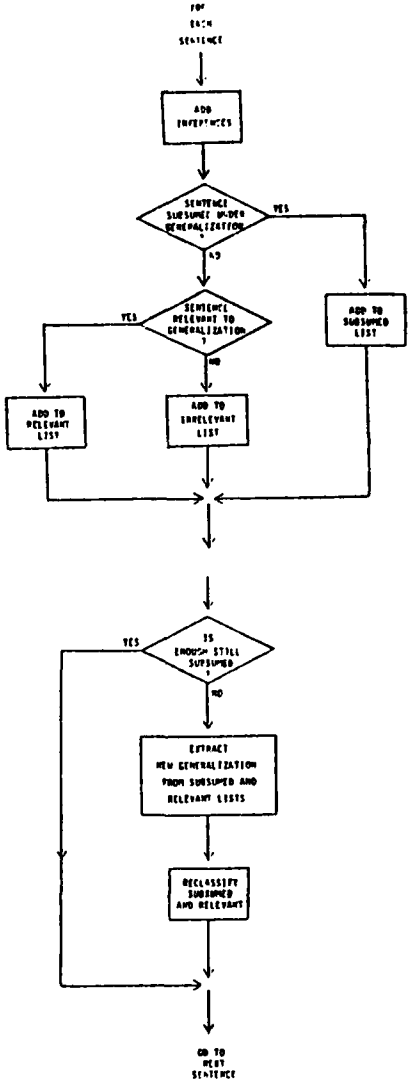


Figure 6. Production Rules for Subsuming Strategy.

passage. If the first sentence is general, a relatively conservative criterion for deciding to revise is used. An example of such a criterion is that if the number of propositions currently either subsumed or related is greater than the number of irrelevant propositions, the candidate main idea is still satisfactory. If the first sentence is not general, a "hair trigger" for revision is used. This criterion can take different forms, for example, (a) if inference had to be done before the sentence could be subsumed, a revision should be done, (b) if the sentence was classified as irrelevant, but contained more than just a few propositions, a revision should be done, (c) if the number of main propositions considered irrelevant is not less than the number of main propositions that have been subsumed, a revision should be done.

These rules are not really completely satisfactory, a point which will be returned to. At this point many different combinations of rules have been tried in the model. The problem is that each subject may have his or her own rules, and these are undoubtedly typically more subtle that the model's rather crude mechanisms would permit. Some useful results with the model have been obtained, however, and will be summarized here.

Comparison of the Simulation and Data

Ratings and protocols

The decisions made by the simulation can be compared to the ratings and protocol data already presented. For the METALS passage, these results are shown in Table 24, which shows the modal ratings, a modal summary of the protocols, and a summary of the model's activities for a particular run using a particular set of strategy options and revision criteria. In the good version, the first sentence is accepted as general, then the other sentences are subsumed, or found irrelevant. There are no revisions. The agreement with the ratings summary is good, and with the protocols, roughly similar. The discrepancy on Sentence 11 suggests that a "nothing new" rule is needed. In the bad version, the results for 2 different hair-trigger rules are shown. The first sentence is generalized in these runs; under a wait-and-see option available in the model, the same generalization would be produced after the second sentence. For the TEST27 run, the simulation attempts to revise the main idea on Sentence 4, but arrives at the same main idea as before. But notice the change in status of this sentence between the two versions. It calls Sentence 6 irrelevant, which contrasts with its subsumed status in the good version. The model revises at Sentence 10 and chooses the intended main idea, but the hair trigger used in this run forces another revison attempt on Sentence 14 because

Table 24

Simulation results for METALS

Good Version

Sent. No.	Rating Mode	Protocol Mode	TEST20
1.	C	A	IS GENERAL (USE CULTURE METAL)
2.	R	S	SUBSUMED
3.	R	S	SUBSUMED
4.	U	R	IRRELEVANT
5.	R	RC	SUBSUMED
6.	R	RC	SUBSUMED
7.	U	I	IRRELEVANT
8.	R	R	SUBSUMED
9.	U	S	IRRELEVANT
10.	R	S	SUBSUMED
11.	U	I,R	SUBSUMED
12.	R	R	SUBSUMED
13.	U	I,R	IRRELEVANT
14.	R	R,S	SUBSUMED (USE CULTURE METAL)

Bad Version

Sent. No.	Rating Mode	Protocol Mode	TEST27	TEST1D
2.	R	G	GENERALIZE (USE CULTURE WEAPON)	GENERALIZE (USE CULTURE WEAPON)
3.	R	G	SUBSUMED	SUBSUMED
4.	R	RC	SUBSUMED DO GEN-ALL,SAME	SUBSUMED
5.	R	RC	SUBSUMED	SUBSUMED
6.	U	RC	IRRELEVANT	IRRELEVANT,GEN-ALL (USE CULTURE METAL)
7.	U	I	IRRELEVANT	IRRELEVANT
8.	R	R	IRRELEVANT	SUBSUMED
9.	U	R	SUBSUMED	RELATED TO SUBSUMED
10.	R	S	IRRELEVANT, NEW:(USE CULTURE METAL)	SUBSUMED
11.	U	R	SUBSUMED	SUBSUMED
12.	R	S	SUBSUMED	SUBSUMED
13.	U	R	IRRELEVANT	IRRELEVANT
14.	U	S	SUBSUMED, DO GEN-ALL,SAME (USE CULTURE METAL)	SUBSUMED (USE CULTURE METAL)

inference was required before subsumption could be done. In the TEST1D run, the model triggers if a non-trivial irrelevant sentence comes in, resulting in a revision to the correct main idea at Sentence 6, and no further revisions.

The TIMEKEEPING passage shows very little difference in responses or reading times between the two versions. A similar effect appears in the simulation results, shown in Table 25. In the bad version, Sentence 2 is generalized to the same main idea as provided in the good version. In both cases the large irrelevant Sentence 5 triggers a revision attempt, but no change. The model, however, does not have the intelligence to engage in the considerable processing that most of the protocol subjects did on this sentence.

In the INSTRUMENTS passage (Table 26), the simulation has a relatively difficult time because the passage sentences require a large number of inferences just to establish the basic coherence of the passage. Again this might explain the reported difficulty of the passage. In the good version, the simulation drops the explicitly presented main idea under the onslaught of repeated sentences that aren't immediately subsumed, and then comes back to the initial main idea at Sentence 10. In the protocols and responses some of this pattern is evident. Due to the complexity of the passage, a satisfactory run for the bad version has not been obtained.

In the CARS passage (Table 27), the simulation keeps the intended main idea in the good version, but repeatedly attempts revisions; the model can not handle the series of apparently irrelevant sentences appearing early in the passage. In the bad version, Sentence 2 is considered general, thanks to the facts in long-term memory, and then Sentences 3 and 4 about luxury cars are subsumed, but then as some subjects did, the simulation abandoned this hypothesis at Sentence 5, and adopted the intended main idea.

Predictions of Reading Times in METALS

Depending on the strategy and the passage version, the simulation may do different amounts of work on some of the sentences, depending on whether a revision is performed on the sentence. Thus, the reading time on the sentences in the two versions should vary in a way related to the amount of work done in the simulation. But recall that in most of the passages, no version effects on reading time appear, and the reading time was predicted very well by superficial predictors, such as the number of words. So the macroprocessing time can not be distinguished from the superficial effects in most cases. But there are version effects in the METALS passage which are related to the simulation's macroprocessing. Using two of the simulation

Table 25

Simulation results for TIMEKEEPING

Good Version

Sent. No.	Rating Mode	Protocol Mode	TEST2A
1.	C	A	IS GENERAL (MOD TKD EX-ACCURATE)
2.	R	S	SUBSUMED
3.	R	R	SUBSUMED
4.	R	S	SUBSUMED
5.	U	R	IRRELEVANT,GEN-ALL, SAME RESULT
6.	R	R	SUBSUMED
7.	R	S	SUBSUMED
8.	R,U	I	SUBSUMED (MOD TKD EX-ACCURATE)

Bad Version

Sent. No.	Rating Mode	Protocol Mode	TEST2B
2.	R	G	GENERALIZE (MOD TKD EX-ACCURATE)
3.	R	R	SUBSUMED
4.	R	S	SUBSUMED
5.	U	R	IRRELEVANT, GEN-ALL, SAME RESULT
6.	R	RC	SUBSUMED
7.	R	S	SUBSUMED
8.	R	R	SUBSUMED (MOD TKD EX-ACCURATE)

Table 26

Simulation results for INSTRUMENTS

Good Version

Sent. No.	Rating Mode	Protocol Mode	TEST3D
1.	C	A	IS GENERAL (CONTROL KBI SOUND-ASPECT)
2.	R	R	IRRELEVANT
3.	R	S	IRRELEVANT,GEN-ALL,NOW SUBSUMED, (POSSESS KBI MECHANISM)
4.	R	S	SUBSUMED
5.	R	S	SUBSUMED
6.	R	S	SUBSUMED
7.	R	S	SUBSUMED
8.	R	S	SUBSUMED
9.	R	S	SUBSUMED
10.	R	R	IRRELEVANT, GEN-ALL
11.	R	R	....

Table 27

Simulation results for CARS

Good Version

Sent. No.	Rating Mode	Protocol Mode	TEST4C
1.	R	S,R	IS GENERAL (SELECT PEOPLE AUTOMOBILE)
2.	R	S	IRRELEVANT, GEN-ALL, SAME RESULT
3.	R	S	SUBSUMED
4.	R	S	IRRELEVANT, GEN-ALL, SAME RESULT
5.	R	S	SUBSUMED
6.	U	I	IRRELEVANT
7.	R	S	SUBSUMED
8.	R	S	IRRELEVANT, GEN-ALL, SAME RESULT
9.	R	S	SUBSUMED
10.	U	I	IRRELEVANT
11.	U	R	IRRELEVANT, GEN-ALL, SAME RESULT
12.	R	S	SUBSUMED
13.	U	R	IRRELEVANT (SELECT PEOPLE AUTOMOBILE)

Bad Version

Sent. No.	Rating Mode	Protocol Mode	TEST4D
2.	C	A	IS GENERAL (POSSESS EIL-C FEATURE)
3.	R	S	SUBSUMED
4.	R	R	SUBSUMED
5.	U	SC	IRRELEVANT, GEN-ALL (SELECT PEOPLE AUTOMOBILE)
6.	U	I	IRRELEVANT
7.	R	S	SUBSUMED
8.	R	S	IRRELEVANT, GEN-ALL,SAME RESULT
9.	R	S	SUBSUMED
10.	R	I	IRRELEVANT
11.	U	R	IRRELEVANT, GEN-ALL, SAME RESULT
12.	R	S	SUBSUMED
13.	R	R	IRRELEVANT (SELECT PEOPLE AUTOMOBILE)

runs shown above, the variable POPRS was defined as the total number of operations performed on propositions by the production rules: the number built, removed from a list, or moved from one list to another. This variable was included in a regression analysis of the mean reading times on each sentence. The predicted and observed times are shown in Figures 5A and 5B. The prediction equation is RT = 1.769 + (.183) WORDS + (1.255) FIRST + (.023) POPRS, which accounts for about 80% of the variance, with all variables contributing significantly at the .05 level. This fit is better than that obtained using the WORDS predictors (Figures 2A and 2B). The good fit is encouraging that the model captures not just the qualitative features of where people revise their main ideas, but also some of the quantitative aspects of the amount of processing performed while reading.

Critique of the Model

The most important failing of the model is that the simple quantity-based revision criteria do not seem to be a very good approach, for two reasons. First, they are a simple sentence-by-sentence process that does not make much use of the overall organization of the passage. That is, the protocol subjects often predicted what they were going to see next, strongly suggesting that they were using a schema for generalization passages. This use of a schema seems to be what enables them to accept sentences that are irrelevant, but that lead up to an instance, such as the first few sentences in the good version of CARS. But the simple quantity-based criteria are unable to handle this problem in a reasonable way. A second problem is that subjects are extremely varied in what they do, as shown emphatically by the protocol results, but as also implied by the large spread in importance ratings and the variety in the main idea responses. It seems rather unlikely that the variety of possible decision rules could be easily represented in terms of different rules for simple quantity comparisons. It would be preferable to capture these differences in terms of either differences in LTM knowledge, or basic process differences, such as differences in the inference or generalization rules used.

But the major contribution of the model is showing that reasonably accurate decisions and main idea responses could be based on rather limited amounts of long-term memory knowledge. For example, Table 28 shows the LTM required for the METALS passage, which has received the most attention in the modelling work. Note that the bulk of the propositions consist simply of ISA relationships, which are those required for the generalization and subsumption rules. The IMPLY propositions are required for the inferences that make implicit propositions explicit, so that the subsumption and generalization rules can use them. These LTM propositions

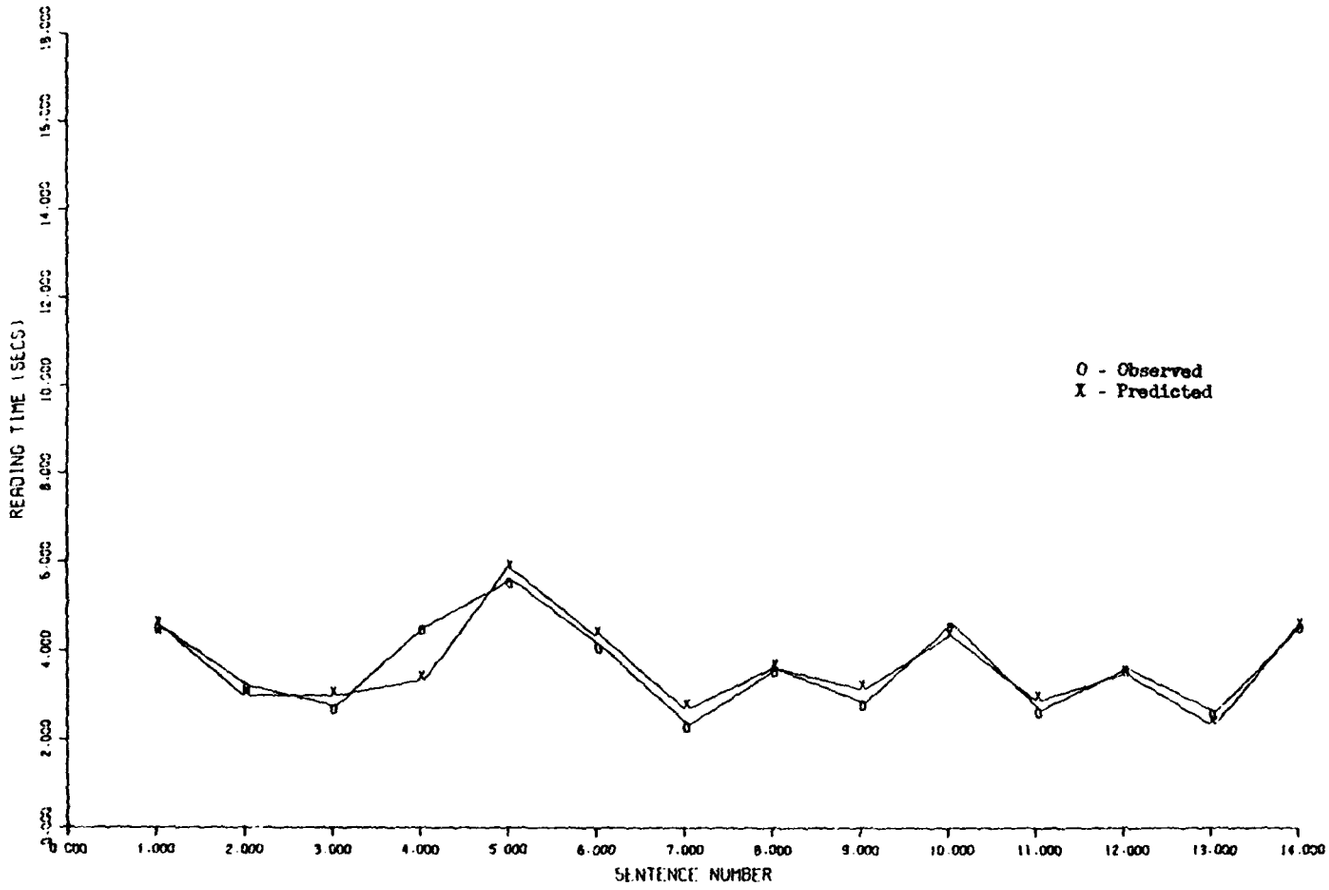


FIGURE 5A. PREDICTED (SIMULATION) VS OBSERVED TIMES FOR METALS. GOOD VERSION

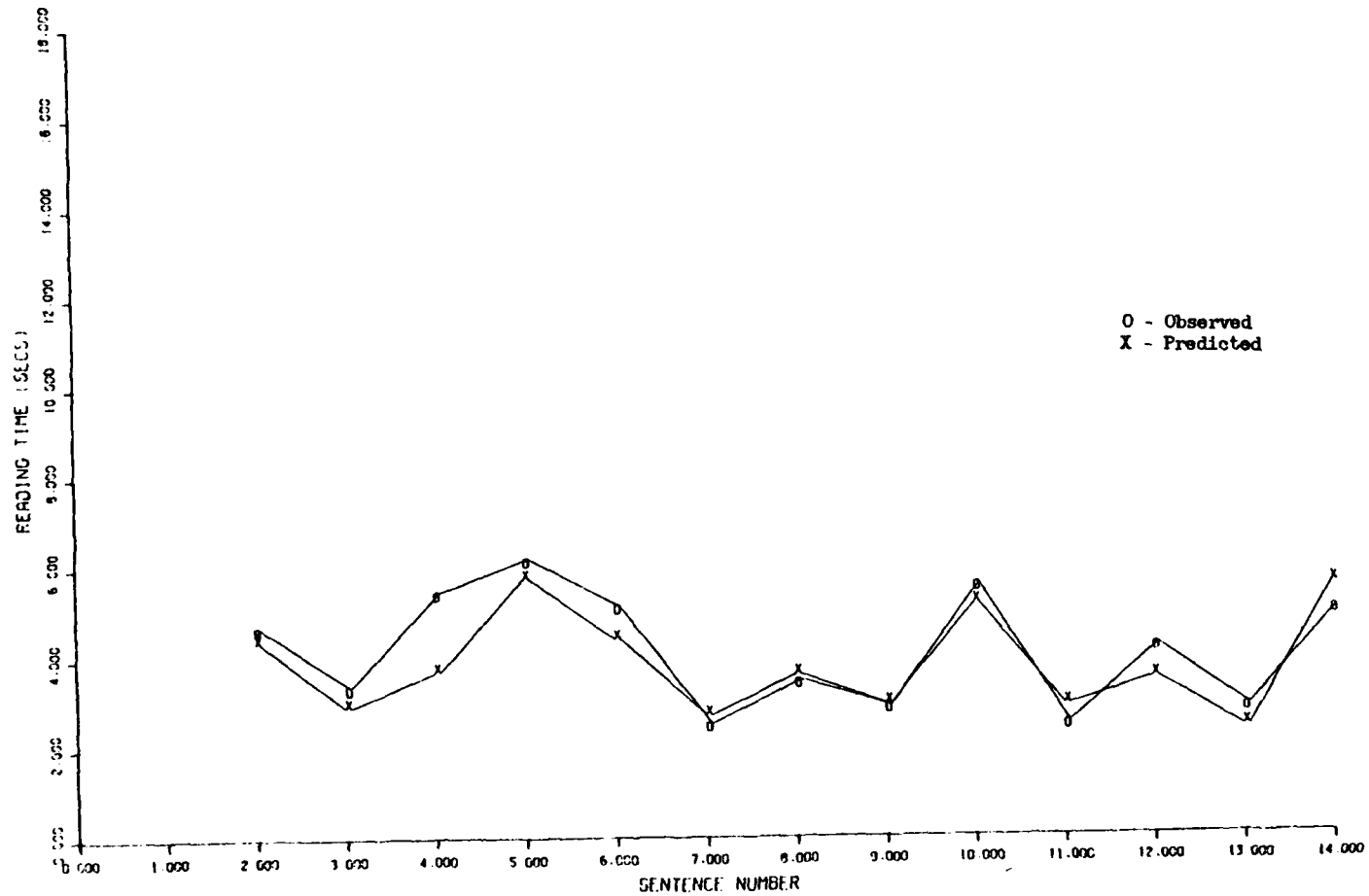


FIGURE 5B. PREDICTED (SIMULATION) VS OBSERVED TIMES FOR METALS, RAD VERSION

Table 28

LTM Used by the Simulation for METALS

```
L1 (ISA HELLENES CULTURE) L2 (ISA GREEKS CULTURE)
L2A (LIVE-IN GREEKS GREECE) L2B (ISA GREECE COUNTRY)
L3 (ISA INCAS CULTURE) L4 (ISA SPANIARDS CULTURE)
L5 (ISA MWCULTURE CULTURE) L6 (ISA $ CULTURE)
L7 (ISA BRONZE METAL) L8 (ISA COPPER METAL)
L9 (ISA GOLD METAL) L10 (ISA ALUMINUM METAL)
L11 (ISA TITANIUM METAL)
L12 (ISA SWORDS WEAPON) L13 (ISA SHIELDS WEAPON)
L14 (ISA WARPLANES WEAPON) L15 (ISA ARTIST CULTURE)
L16 (ISA PERSON $)
INF1  (IMPLY (BOTH (SAME-AS *Z1 *Z2) (*Z3 *Z2 *Z4))
             (*Z3 *Z1 *Z4))
INF1A (IMPLY (BOTH (SAME-AS *Z1 *Z2) (*Z3 *Z1 *Z4))
             (*Z3 *Z2 *Z4))
GK1 (IMPLY (VALUEV *Z1 *Z2) (USE *Z1 *Z2))
GK2 (IMPLY (MOD *Z1 POPULAR) (USE $ *Z1))
GK3 (IMPLY (ESSENTIAL-FOR *Z1 *Z2) (USE $ *Z1))
GK4 (IMPLY (WANT *Z1 *Z2) (USE *Z1 *Z2))
GK5 (IMPLY (INVADE *Z1 *Z2) (USE *Z1 WEAPON))
GK6 (IMPLY (CONQUER *Z1 *Z2) (USE *Z1 WEAPON))
GK7 (IMPLY (CONQUER *Z1 *Z2) (INVADE *Z1 *Z2))
GK8 (IMPLY (CUT-THROUGH *Z1 *Z2) (SUPERIOR *Z1 *Z2))
GK9 (IMPLY (BOTH (USE *Z1 *Z2) (MADE-OF *Z2 *Z3))
           (USE *Z1 *Z3))
GK10 (IMPLY (BOTH (LIVE-IN *Z1 *Z2) (*Z3 *Z4 *Z2))
            (*Z3 *Z4 *Z1))
GK11 (IMPLY (BOTH (WANT *Z1 *Z2) (BELONG-TO *Z2 *Z3))
            (WANT *Z1 *Z3))
GK12 (IMPLY (BOTH (USE *Z1 *Z2) (IN *Z3 *Z2))
            (USE *Z1 *Z3))
GK13 (IMPLY (BOTH (INVADE *Z1 *Z2)(BEAT *Z1 *Z2))
            (CONQUER *Z1 *Z2))
GK13A (IMPLY (BEAT *Z1 *Z2) (CONQUER *Z1 *Z2))
GK15 (COMPIMPLY (ALL5 (USE *Z1 *Z2) (ISA *Z2 WEAPON)
                      (USE *Z3 *Z4) (ISA *Z4 WEAPON)
                      (SUPERIOR *Z2 *Z4))
                (BEAT *Z1 *Z3))
GK16 (IMPLY (ALL3 (INVADE *Z1 *Z2) (ISA *Z2 COUNTRY)
                  (LIVE-IN *Z3 *Z2))
            (CONQUER *Z1 *Z3))
```

are a rather small subset of the possible general knowledge related to this passage.

This conclusion ties back to earlier results on the abstraction task (Kieras, 1980). In picking and producing topical or thematic information from technical material, people can make use of the semantic content, even though they do not understand the material deeply at all. A good example from these results is the comments of one of the protocol subjects who in the timekeeping passage said of Sentence 7: "I don't know what a hydrogen maser is, and I don't know what a picosecond is, but it is obviously a clock that is extremely accurate." Like this subject, the model also has an extremely limited understanding of the material, but it can produce main ideas and judge sentence relevance with only this very superficial knowledge. That only "shallow semantics" might suffice for a great deal of macrostructure processing is a useful theoretical conclusion.

Reference Notes

1. Kieras, D. E. Abstracting main ideas from technical prose: A preliminary study of six passages. Technical Report No. 5, University of Arizona, August, 1980.
2. Bovair, S., & Kieras, D. E. A Guide to Propositional Analysis for Research on Technical Prose. Technical Report No. 8, University of Arizona, July, 1981.
3. Turner, A., & Greene, E. The construction and use of a propositional text base. Institute for the Study of Intellectual Behavior, Technical Report No. 63. University of Colorado, April, 1977.
4. Kieras, D. E. The relation of topics and themes in naturally occurring technical paragraphs. Technical Report No. 1, University of Arizona, January, 1979.

References

Clements, P. The effects of staging on recall from prose. In R. O. Freedle (Ed.), New directions in discourse processing. Norwood, New Jersey: Ablex Publishing Corporation, 1979.

Kieras, D. E. Good and bad structure in simple paragraphs: Effects on apparent theme, reading time, and recall. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 13-28.

Kieras, D. E. Doing it the vendor's way: Running multiple subjects in reading experiments using Data General's Diskette Operating System. Behavior Research Methods and Instrumentation, 1979, 11, 221-224.

Kieras, D. E. Initial mention as a signal to thematic content in technical passages. Memory & Cognition, 1980, 8, 345-353.

Kieras, D. E. The role of major referents and sentence topics in the construction of passage macrostructure. Discourse Processes, 1981, 4, 1-15.

Kieras, D. E. Component processes in the comprehension of simple prose. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 1-23.

Kintsch, W. The representation of meaning in memory. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1974.

Kintsch, W., & van Dijk, T. A. Toward a model of discourse comprehension and production. Psychological Review, 1978, 85, 363-394.

Kozminsky, E. Altering comprehension: The effect of biasing

titles on text comprehension. Memory & Cognition, 1977, 5, 482-490.

Perfetti, C. A., & Goldman, S. R. Thematization and sentence retrieval. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 70-79.

Perfetti, C. A., & Goldman, S. R. Discourse functions of thematization and topicalization. Journal of Psycholinguistic Research, 1975, 4, 257-271.

van Dijk, T. A. Text and context. London: Longman, 1977. (a)

van Dijk, T. A. Semantic macro-structures and knowledge frames in discourse comprehension. In M. Just & P. Carpenter (Eds), Cognitive processes in comprehension. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977. (b)

van Dijk, T. A. Relevance assignment in discourse comprehension. Discourse Processes, 1979, 2, 113-126.

van Dijk, T. A. Macrostructures. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Meryl S. Baker
NPRDC
Code P309
San Diego, CA 92152

1 Dr. Robert Breaux
Code N-711
NAVTRAEQUIPCEN
Orlando, FL 32813

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Jim Hollan
Code 304
Navy Personnel R & D Center
San Diego, CA 92152

1 CDR Charles W. Hutchins
Naval Air Systems Command Hq
AIR-340F
Navy Department
Washington, DC 20361

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

Navy

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

Navy

1 Special Asst. for Education and Training (OP-01E)
Rm. 2705 Arlington Annex
Washington, DC 20370

1 Office of the Chief of Naval Operations
Research Development & Studies Branch (OP-115)
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
Selection and Training Research Division
Human Performance Sciences Dept.
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Dr. Bernard Rimland (03B)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland, Director
Research, Development, Test & Evaluation
N-5
Naval Education and Training Command
NAS, Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group (TAEG)
Dept. of the Navy
Orlando, FL 32813

Navy

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Roger Weissinger-Baylon
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr John H. Wolfe
Code P310
U. S. Navy Personnel Research and Development Center
San Diego, CA 92152

Army

1 Technical Director
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. James Baker
Systems Manning Technical Area
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

1 Dr. Beatrice J. Farr
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 DR. FRANK J. HARRIS
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Col Frank Hart
Army Research Institute for the Behavioral & Social Sciences
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Army

1 Dr. Robert Sasmor
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

2 3700 TCHTW/TTGH Stop 32
Sheppard AFB, TX 76311

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Chief, Psychological Reserch Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

Other DoD

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civil Govt

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 William J. McLaurin
66610 Howie Court
Camp Springs, MD 20031

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Joseph Psotka
National Institute of Education
1200 19th St. NW
Washington,DC 20208

1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Anderson, Thomas H., Ph.D.
Center for the Study of Reading
174 Children's Research Center
51 Gerty Drive
Champiagn, IL 61820

1 Dr. John Annett
Department of Psychology
University of Warwick
Coventry CV4 7AL
ENGLAND

1 DR. MICHAEL ATWOOD
SCIENCE APPLICATIONS INSTITUTE
40 DENVER TECH. CENTER WEST
7935 E. PRENTICE AVENUE
ENGLEWOOD, CO 80110

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Patricia Baggett
Department of Psychology
University of Colorado
Boulder, CO 80309

1 Mr Avron Barr
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Liaison Scientists
Office of Naval Research,
Branch Office , London
Box 39 FPO New York 09510

Non Govt

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80309

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 Dr. Bruce Buchanan
Department of Computer Science
Stanford University
Stanford, CA 94305

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94305

Non Govt

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Drawer Q
Santa Barbara, CA 93102

1 LCOL J. C. Eggenberger
DIRECTORATE OF PERSONNEL APPLIED RESEARC
NATIONAL DEFENCE HQ
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A OK2

1 Dr. Ed Feigenbaum
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Mr. Wallace Feurzeig
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

Non Govt

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2E9

1 Dr. R. Edward Geiselman
Department of Psychology
University of California
Los Angeles, CA 90024

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Daniel Gopher
Industrial & Management Engineering
Technion-Israel Institute of Technology
Haifa
ISRAEL

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Harold Hawkins
Department of Psychology
University of Oregon
Eugene OR 97403

1 Dr. Barbara Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

Non Govt

1 Dr. Frederick Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. James R. Hoffman
Department of Psychology
University of Delaware
Newark, DE 19711

1 Dr. Kristina Hooper
Clark Kerr Hall
University of California
Santa Cruz, CA 95060

1 Glenda Greenwald, Ed.
"Human Intelligence Newsletter"
P. O. Box 1163
Birmingham, MI 48012

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Ed Hutchins
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Dr. Marcy Lansman
Department of Psychology, NI 25
University of Washington
Seattle, WA 98195

Non Govt

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

1 Mr. Merl Malehorn
Dept. of Navy
Chief of Naval Operations
OP-113
Washington, DC 20350

1 Dr. Erik McWilliams
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550

1 Dr. Mark Miller
TI Computer Science Lab
C/O 2824 Winterplace Circle
Plano, TX 75075

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. Donald A Norman
Dept. of Psychology C-009
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Seymour A. Papert
Massachusetts Institute of Technology
Artificial Intelligence Lab
545 Technology Square
Cambridge, MA 02139

Non Govt

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barabara, CA 93106

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 Dr. Martha Polson
Department of Psychology
Campus Box 346
University of Colorado
Boulder, CO 80309

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80309

1 Dr. Steven E. Poltrock
Department of Psychology
University of Denver
Denver,CO 80208

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Lauren Resnick
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

Non Govt

1 Mary Riley
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

1 Dr. Alan Schoenfeld
Department of Mathematics
Hamilton College
Clinton, NY 13323

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Committee on Cognitive Research
% Dr. Lonnie R. Sherrod
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Alexander W. Siegel
Department of Psychology
SR-1
University of Houston
Houston, TX 77004

Non Govt

1 Robert S. Siegler
Associate Professor
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

1 Dr. Edward E. Smith
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 Dr. Thomas G. Sticht
Director, Basic Skills Division
HUMRRO
300 N. Washington Street
Alexandria,VA 22314

1 David E. Stone, Ph.D.
Hazeltine Corporation
7680 Old Springhouse Road
McLean, VA 22102

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

Non Govt

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. John Thomas
IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

1 DR. PERRY THORNDYKE
THE RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CA 90406

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Benton J. Underwood
Dept. of Psychology
Northwestern University
Evanston, IL 60201

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 DR. GERSHON WELTMAN
PERCEPTRONICS INC.
6271 VARIEL AVE.
WOODLAND HILLS, CA 91367

1 Dr. Keith T. Wescourt
Information Sciences Dept.
The Rand Corporation
1700 Main St.